

Tilburg University

Care to explain?

de Groot, Aviva

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
de Groot, A. (2023). *Care to explain? A critical epistemic in/justice based analysis of legal explanation obligations and ideals for 'AI'-infused times.* [Doctoral Thesis, Tilburg University].

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Aviva de Groot

Care to explain?

A critical epistemic in/justice-based
analysis of legal explanation obligations and ideals
for 'AI'-infused times

Care to explain?

A critical epistemic in/justice-based analysis
of legal explanation obligations and ideals for ‘AI’-infused times

Aviva de Groot

Care to explain?

A critical epistemic in/justice-based analysis
of legal explanation obligations and ideals for ‘AI’-infused times

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University op
gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het openbaar
te verdedigen ten overstaan van een door het college voor promoties aangewezen
commissie in de aula van de Universiteit op 12 mei 2023 om 13.30 uur

door

Aviva de Groot,

geboren te Amsterdam

Promotores: Professor Dr. R.E. Leenes (Tilburg University)
Professor Dr. N.N. Purtova (Utrecht University)

Promotiecommissie: Dr. H. Felzmann (University of Galway)
Professor Dr. B.J. Koops (Tilburg University))
Professor Dr. S. Ranchordás (University of Tilburg)
Professor Dr. D. L. Willems (Amsterdam UMC / University
of Amsterdam)

This research was funded in part by a grant from KPN

ISBN: 978-94-93315-50-1

Gedrukt door: Proefschrift-AIO.nl

© Aviva de Groot, The Netherlands. All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means without permission of the author. Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd, in enige vorm of op enige wijze, zonder voorafgaande schriftelijke toestemming van de auteur.

The cover design traveled with the research process and involved Friso, Aviva, Maarten, and finally and decisively Joost (THANK YOU!) at different stages.

Acknowledgements

Lucky me, I mused when I started this research project. I have a software engineering partner to query. I was hired to wrestle a specific legal demand for the explanation of automated decisions and I imagined we could run a pet algorithmic project to test my hypotheses on. I guess that worked. The design and development of an automated household decision aid led to productive and increasingly serious arguments about the extent of mutual understanding that we could, and should, be able to reach. There were times I wanted to give up on the whole mutual understanding thing. Unwilling to simply ‘trust’ however, I requested that I was at least met with a responsible display of trustworthiness.¹ But what is that?

The development now rests but not the thinking. Some label the completion of a PhD thesis as a process, some call it a journey. I’ll call it a phase in life-long learning. This book is a snapshot of a work in progress. It resonates with thought and experience from before, and equips me for learning that is still to come.

Along the way, ‘before’ also gained specific meaning for myself (and all my colleagues.) Before the 2020 Corona pandemic I got to run around, cross paths with differently disciplined researchers and pester them with my thoughts and questions, organize discussions, and be with people ‘on the ground’ whose experience my thesis aims to serve. As hard as learning can be, I counted myself lucky to get to do this on a salary. When Covid ‘hit’ the Netherlands these interactions fell away and a more isolated process began. At the same time, the acute proliferation of crisis politics fit the theme of what I, by then, was doing: questioning the novelty of problems presented as consequences of ‘crises,’ and studying the politics of justification. The experiences culminated in a somewhat cynical 2021 conference presentation on ‘badly explained pandemic politics’ – which I had put to rhyme and mixed with soundbites of Ramones, RATM, and Richard Hell (“Betrayal takes two / who did it to who”).

By then Ronald, my main supervisor and in charge of the department in those hectic times, trusted me (I think) to come around with something useful and understandable to say. His early rebuttals along the lines of “so then convince us that ‘x’ matters” made me keen to deliver. Ronald, having you on board, especially in the challenging last phase has made all the difference. Nadya, thank you having the confidence to let me run off and return with wheelbarrows of hard-to-parse argument built on alien bodies of literature, and patiently pointing out where humane readable translation was needed. It can’t have been easy!

1 For who wants to know: the tool decides when and how to turn, flip, or flop (“Please. Rotate x, y, z”) the mattress and I still don’t trust that it serves the demands of our different size and weight. We agree that it is hard to settle on a certain amount of informedness on mattresses.

I want to thank inspirators Anja and Huib who were part of my life at different times and who, in their own and frequently impossible ways, showed me what fighting injustice entails. I wish you had lived to the defense to query me and my opponents unannounced, as I believe you would have done. Never questioning your own motivation while you went the lengths you did, you sometimes had little patience with those who questioned you. I guess in retrospect, I'm trying to help out there.

I have come to understand how it was no coincidence that these persons would simply treat me as an active partner in work I still needed to learn how to do (take care of people, sue the State). I was raised by my grandmothers' embodied survival and my mother's untheorized and caring brand of independence. If they shine through in what motivates me, it is because I am here for them. Thank you both.

Thank you also to those colleagues and teachers who I met at work and late-in-life legal studies, and whose casual projections of me doing a PhD seeded my own belief that I could do such a thing – Ine, Bart, Bojana, Tarlach, Kristina, Eduard. During the PhD years, themselves, I'm grateful to have enjoyed the company of friendly colleagues from different places with whom I could effortlessly switch between shoptalk, life talk, jokes, or loud concert attendances: Ine, Tineke, Bojana, Shaz, Gijs, Maranke, Jenneke and many others. Serendipitous conversations (and the rare planned follow-up) with Robin and Bert-Jaap have been vital to my thinking, and thank you Jenneke for locating crucial texts after the pandemic closing of medical libraries (and other crucial times ;) Tineke, your literal and virtual hospitality were cherished antidotes to the more exhausting dimensions of doing this PhD. I hope to trade many more train hours for food-and-conversation hours. To all @ TILT: please feel included in these acknowledgements, you all make this a special place to work. I'm happy to write 'make' and not 'made,' and look forward to the next few years of continued work in Tilburg – and beyond. Born out of necessity in the pandemic, many gatherings are now accessible online to allow for remote participation. I'm grateful for how this allowed me to meet the joyfully witty ACEPS philosophers who inspire me to continue on this research path. Veli, Akanimo, Caitlin and Abe – here's to more co-laborations between our institutes.

My family and (other) friends have been indispensable. I'm grateful that none of you asked that I stop this nonsense, and instead humored me with food for thought, body, and spirit. Here's to more time to spend well together. Zus, *thank you* for dragging me outside to exercise, and for JoJo and the 3 heroes. To my daughters Zazie & Loes, to Bas, and to you-who-I-will-enjoy-to-make-a-fool-of-myself-for-from-the-day-you-are-born: your presences are the heartbeat of my life and I would not be able to make sense of it without you.

M, how on earth can I thank you properly? I've hardly a clue of the cunningness of your care. For what I do know¹ I am immodestly grateful, and I embrace you.

Last but not least: Sheena, if it weren't for you ..ehm, well I guess there would be no product to contain these printed words. Not sure it matters, so best rock on!

1 The sound, the food, the thought ..! Here's to jumping to unlikely beats, and to pandemic dino's.

Table of contents

Acknowledgements.....	VIII
-----------------------	------

1	Introduction.....	1
1.1	The right (to) explanation debate.....	1
1.1.1	A j got stuck in the system	1
1.1.2	Explanation as a fundamental condition for humane treatment	2
1.1.3	What makes a humane explainer?	4
1.1.4	Research objective	6
1.2	Methodological approach and structure	6
1.2.1	An iterative research process	6
1.2.1.1	Between moving targets and sitting ducks	6
1.2.1.2	Disconnecting from automation	10
1.2.2	A critical analysis of legal explanation rules	11
1.2.2.1	Outwith law	11
1.2.2.2	The critical angle of choice: epistemic injustice and justice literature	13
1.2.2.3	Theoretical application	14
1.2.3	Approach of two regulated explanation domains.....	15
1.2.3.1	Selection and focus	15
1.2.3.2	Decisional whats, who's, hows, and explanation rules: descriptive parts of the domain chapters.....	17
1.2.3.3	Qualification of the explanation rules in terms of the thesis's theoretical development.....	18
1.2.4	Contribution of the thesis.....	18
1.2.5	Thesis structure.....	19
2	New crisis, old problems: sourcing research questions from a tension	23
2.1	Explanation rules and aims: a very short introduction	23
2.1.1	Regulated explanation in situations of substantive power and information inequality.....	23
2.1.2	Clarification, explanation, justification: instrumental aims of explanation rules.....	25
2.1.3	Dignitarian aims and arguments.....	27
2.2	Perceptions on a right in crisis	32
2.2.1	Decisions are not made here anymore: a crisis of the individual right.....	32
2.2.2	Eroding possibilities for meaningful subject participation	35
2.2.3	Explainers are challenged to fulfill their duties	37
2.2.4	Challenges to typical explanation values, or: denial of crisis	41
2.3	Problems of knowledge: four provocations of 'new'	44
2.3.1	Sense making on individual levels as a lucky draw	45
2.3.2	Idealistic assumptions about knowledgeable participation	46
2.3.3	Who respects whose human ability to reason?	48

2.3.4	Explainers are always challenged, whether they know it or not	50
2.4	Are our explanation rules in need of justification?	53
2.4.1	Two meaningful applications of ‘meaningful’	53
2.4.1.1	Meaningful information positions	53
2.4.1.2	Meaningful explanation rules?	55
2.4.2	Research question and sub questions	58
2.5	Chapter 2 in a nutshell.....	59
3	Meaningful information positioning: an application of epistemic injustice and justice theory to explanation regulation.....	61
3.1	Re-idealizing explanation in recognition of non-ideal theory.....	61
3.1.1	Arguments for a ‘negatively informed’ theoretical engagement	61
3.1.1.1	Note on literature	63
3.1.2	Application of the consulted theories	65
3.2	Injustice and justice in knowledge practices	66
3.2.1	Key concepts: knowledge, justice, and truth	66
3.2.1.1	Knowledge as a product of sociality	66
3.2.1.2	Justice: respecting and promoting fundamental equality of inter-dependent humans	68
3.2.1.3	Truth: the merits of chasing non-oppressive representations.....	69
3.2.2	Bad: direct harms as an effect of misconstruction and misuse of epistemic authority	72
3.2.3	Murky: perpetuation of wrongs in and through shared knowledge spheres.....	76
3.2.3.1	The importance and influence of science as a knowledge sphere (revisited)	76
3.2.3.2	Low maintenance oppression in common knowledge spheres	78
3.2.4	Good: societal and institutional promotion of preventative and corrective labor	80
3.2.4.1	Caring about trustworthiness: where methods and disposition meet	80
3.2.4.2	Doubt, resistance, and a fair distribution of burdens.....	83
3.2.5	Takeaways for ‘explanation’	87
3.3	Aiming for justice in explanation practice.....	89
3.3.1	Meaningful information positions: prerequisite and aim of responsible explanation practices	89
3.3.1.1	Recap: the quest for meaningful information positions.....	89
3.3.1.2	Explanation as an interactive, testimonial practice	90
3.3.1.3	Interactional justice demands	93
3.3.1.4	Beyond statements of reasons: material support for progressive development.....	95
3.3.2	Takeaways for explanation rules	96
3.3.3	Meaningful information position-ing: in rules	97
3.3.3.1	Rules say, rules do	97
3.3.3.2	Saying what rules should to: duties of care for regulated explanation practices.....	98
3.4	Chapter 3 in a nutshell.....	101

4	Meaningful information positioning and legal administrative explanation rules	105
4.1	Introduction	105
4.1.1	Function and value of the administrative domain study	105
4.1.2	Research and reporting choices	107
4.2	Individual administrative decision making: a functional characterization	109
4.2.1	Administrative decisions: “a fact of life” for all Dutch citizens (the ‘what’)...	109
4.2.1.1	Powers of administrative bodies	109
4.2.1.2	Administrative bodies and political climates: realistic expectations	113
4.2.1.3	How all this is legal: the law that regulates powers of, for, and against the administration	116
4.2.1.4	Did Dutch Administrative Law’s main injustice valve dysfunction? The ‘hardship clause’ discussion.	119
4.2.2	Civil servants as interactional partners: expectations and concerns (‘the who’)	122
4.2.3	Individual Administrative decision making: basic norms and instructions (‘the how’)	127
4.2.3.1	Necessary knowledge about relevant facts	128
4.2.3.2	Necessary knowledge about interests to balance: pro’s and cons of the ‘specialty’ principle	131
4.2.3.3	The Dutch Article 22 exception clause: a codified loss of functionality?...	132
4.2.3.4	Imagining categories of people: some words on legal reasoning	134
4.2.3.5	Case illustration 1: ‘tailor made’ decision making in absence of reasoned categories	136
4.2.3.6	Case illustration 2: administrative truths & real life effects of phantom vehicle license registrations	139
4.3	“A decision needs to be supported by a proper motivation”: explanation in the General Administrative Law Act (‘Awb’)	144
4.3.1	Introduction: scope and focus of the Awb’s explanation paradigm	144
4.3.1.1	Dutch Administrative law’s restricted definition of ‘decision’	144
4.3.1.2	The internal review procedure: the entry (and exit) level for elaborate explanations	146
4.3.2	The codification of the principle of Proper Motivation	148
4.3.2.1	“Just 9 (!) words”: the “nihilistic” codification of the principle of motivation	148
4.3.2.2	Can good decisions hide behind bad reasons (and should they be allowed to)? A salient codification discussion	149
4.3.2.3	Proper reasons: “it shouldn’t be much of a burden”	151
4.3.3	The Awb’s main explanation rules: objectives and critical perspectives	152
4.3.3.1	Knowable and insightful	152
4.3.3.2	Reasons in policy rules and the (lesser) motivational burden	153
4.3.3.3	External insightfulness v. a relational understanding of ‘proper’ reasons...	155

4.3.3.4	Decisions ‘as’ arguments v. decisions ‘and’ arguments.....	156
4.3.3.5	Case illustration 3: experiments with informal review procedures.....	157
4.4.	The Awb’s explanation governance in terms of the modeled duties of explanation care.....	161
4.4.1	Introduction: analysis and reporting structure	161
4.4.2	Element one: investigating explainer authority	162
4.4.2.1	General recognition for element one: expectations with regard to civil servants’ critical engagement.....	162
4.4.2.2	Discretionary space for decision makers: what guidance for principled engagement?	164
4.4.2.3	What should reasons represent? The Awb relations of due diligence and motivation	166
4.4.2.4	Reasons on demand: hierarchical relations of decision makers.....	167
4.4.3	Element two: engaging with the social-epistemic positions of explainees.....	167
4.4.3.1	General recognition for element two: the need to look beyond the individual case	167
4.4.3.2	Information positions of decision subjects: the need to go beyond the ‘complexity’ argument.....	168
4.4.3.3	What should reasons represent? A case for explainee-insightfulness	170
4.4.3.4	Making discretionary space work for explainees, with explainees.....	170
4.4.4	Element three: practicing interactional justice	171
4.4.4.1	General recognition for element three: a momentum for legitimacy?	171
4.4.4.2	Responsive information needs of decision subjects: pointed out, seen, but not captured	172
4.4.4.3	Discretionary space for decision makers: a case for formality (revisited)..	172
4.4.4.4	What should reasons represent?	174
4.4.4.5	Reasons on demand in relation to the principle of motivation’s ‘compensation’ function	174
4.4.5	Element four: creating records	175
4.5	Chapter 4 in a nutshell.....	176
5	Meaningful information positioning and legal explanation rules for General (medical) Practice	181
5.1	Introduction	181
5.1.1	Function and value of the General (medical) Practice domain study	181
5.1.2	Research and reporting choices.....	182
5.2	General Practice decision making in The Netherlands: a functional characterization	185
5.2.1	Elementary GP decision making: diagnosis (‘the what’)	185
5.2.1.1	Introduction.....	185
5.2.1.2	The sociality of diagnosis.....	186
5.2.2	Historical and contemporary roles of Dutch GPs (‘the who’).....	190

5.2.2.1	Grappling with paternalism in the formative years of medical ‘people specialists’	190
5.2.2.2	General Practitioners today: personal and political relations to maintain and resist	193
5.2.2.3	Medical expert, communicator, collaborator, leader, health advocate, scholar, professional: GPs in the national training curriculum	194
5.2.3	EBM and SDM: making two core paradigms work for GP practice (‘the how’)	195
5.2.3.1	Introduction	195
5.2.3.2	Evidence Based Medicine: merits, critiques, and alternatives	197
5.2.3.3	Harmonizing doctor and patient knowledge in Shared Decision Making: aims and challenges	200
5.3	‘Information duties’ in the Medical Treatment Agreement Act (‘WGBO’)	204
5.3.1	Introduction	204
5.3.2	The WGBO’s explanation regime in the larger governance landscape	206
5.3.2.1	Power sharing conundrums	206
5.3.2.2	‘Information duties’ as contractual obligations in WGBO: ambition and reception	208
5.3.2.3	Legal text and recent amendments	211
5.3.3	Protecting patients from ‘the risk that they cannot self-determine’	212
5.3.3.1	Introduction	212
5.3.3.2	Different ethical approaches of informed autonomous decision making: a very short introduction	212
5.3.3.3	Patient, person, contractual partner: choices in law about whose ‘free choice’ to serve	216
5.3.3.4	Exceptions to the legal duty to inform: choices and guidance in the WGBO	218
5.3.4	Supporting patients’ decisional capabilities: the legal uptake of SDM	220
5.3.4.1	Introduction	220
5.3.4.2	Added SDM obligations, still a weak promotion of the ‘right’ informed consent?	220
5.3.4.3	How social inequality exacerbates informational inequality: problems with complaints procedures	223
5.3.4.4	Discussing risk: when making conversation turns into making preferences	224
5.4	The WGBO’s explanation governance in terms of the modeled duties of explanation care	225
5.4.1	Introduction	225
5.4.2	Element one: investigating explainer authority	226
5.4.2.1	Recognition of element one: problems in the absence of non-binding obligations for self-reflection	226
5.4.2.2	Promotion of critical and justifiable practices: recognition in the non-legal fields	228

5.4.2.3	Promotion of justifiable knowledge authority: law as the least ambitious norm setter?	229
5.4.3	Element 2: engaging with the social-epistemic positions of explainees.....	232
5.4.3.1	The need to go beyond understanding ‘on behalf of’ patients: support for element two in ethical and professional discourse	232
5.4.3.2	Understanding the process from patients’ point of view: what guidance from law?	234
5.4.4	Element three: practicing interactional justice	236
5.4.4.1	Governance of social-epistemic inequality in SDM: a critical need.....	236
5.4.4.2	Social-epistemic interaction in the WGB0: little support or.. potentially undermining?	238
5.4.5	Element 4: creating records.....	240
5.5	Chapter 5 in a nutshell.....	241
6	Toward care-ful legal explanation regulation: lessons from the present, for the present.....	247
6.1	Looking back, looking forward	247
6.1.1	Drawing lessons from the domain studies	247
6.1.2	Approach and structure	249
6.1.2.1	A note on literature.....	249
6.2	Investigating explainer authority (element one).....	251
6.2.1	General observation	251
6.2.2	Observations from the Administrative domain	251
6.2.2.1	Resistance to progressive understanding	251
6.2.2.2	Unreasoned hardship	253
6.2.2.3	Due diligence and motivation: a case for (more) interdependence	254
6.2.2.4	Further development of the principle of motivation: external v. relational insightfulness	256
6.2.2.5	The focus on meaningful reviewers, problematized (introduction)	256
6.2.3	Observations from the GP domain	258
6.2.3.1	Bad knowledge practices flourish in absence of explanation obligations and relations (it should not need repeating).....	258
6.2.3.2	The technological complexity argument.....	260
6.2.3.3	What’s keeping law from ‘realizing the best possible care’?	262
6.2.4	Suggested emphasis for element one.....	263
6.3	Engaging with the social-epistemic positions of explainees (element two). 264	
6.3.1	General observation	264
6.3.2	Illustrations from the administrative domain.....	264
6.3.2.1	Does the “European unease” with automation need a progressive update?.....	264
6.3.2.2	Don’t ignore the messenger: lessons to learn from the ‘analog’ Administrative domain	265
6.3.2.3	Integrative approaches.....	267

6.3.2.4	Using AIAs to foster explanatory clues for responsible participation	268
6.3.3	Observations from the GP domain	270
6.3.3.1	Times are different now?	270
6.3.3.2	Professional norms of engagement to explicate, boost, & codify.....	271
6.4	Practicing interactional justice (element three)	272
6.4.1	General observation(s).....	272
6.4.2	Observations from the administrative domain.....	273
6.4.2.1	(Finally) explicating principles: duties of care as ‘rules of engagement’ ...	273
6.4.2.2	One more time with feeling: pitfalls of the focus on review	275
6.4.3	Observations from the GP domain	276
6.4.3.1	Be careful what you list for	276
6.4.3.2	The role of law in ‘realizing the best possible care’: arguments for a progressive uptake of professional norms in AI-informed times	278
6.4.3.3	Honesty in the form of sharing conscientious concerns	279
6.5	Creating records (element four).....	281
6.5.1	General observation	281
6.5.2	Observations from the Administrative domain	281
6.5.3	Observations from the GP domain	283
6.6	Mobilizing observations for AI-infused times.....	284
6.6.1	One	285
6.6.2	Two.....	289
6.6.3	Three	292
6.6.4	Four	294
7	The dissertation in a nutshell	297
7.1	De-idealizing and re-idealizing explanation rules	297
7.2	A tale of two domain studies.....	299
7.3	An argument for care-ful progress of present and future explanation regulation	301
	Bibliography	305

1 Introduction

1.1 The right (to) explanation debate

1.1.1 A j got stuck in the system

“Thank you for your email (and your patience). You are totally right, precision is of the essence. I am afraid a j got stuck in the system. I have now corrected this. Our apologies.”²

If the newspaper employee who sent me this message had taken the historical context of the situation into account, they might have thought twice about blaming the system for this particular error. But they did not, and I was landed with the perfect opening statement for a thesis on responsible explainer behavior in automated times.

The correction I had applied for concerned the misspelling of a Holocaust survivor’s surname in a news item about Amsterdam’s new Holocaust Names Memorial.³ The names of 102.000 proverbial ‘j’s, Dutch Jews, and of 220 Sinti and Roma who were murdered in the Second World War are carved in stone.

A stamped ‘j’ on the Jews’ ID cards helped to select them for deportation and from there, to destruction. Fueled by the Nazi’s eagerness for innovative bureaucratic support across their areas of operation, the administrative sorting of people became an increasingly sophisticated automated process.⁴ In The Netherlands and elsewhere, IBM’s punch card technology developed at pace during the larger WWII era.⁵ IBM’s machines could cope with ever broadening ranges of personal data and combine these in ever ‘smarter’ ways to enable ever finer grained handling of people. Combinations of attributes, among which the ‘j’s on the Dutch identity cards translated to patterns of holes in IBM’s punch cards. Fatal combinations of attributes determined any individual ‘j’s route through the system right down to their final destination. And so, between the newspaper employee’s acknowledgement of my claim, and the resolution with apologies, sits a statement that reverberates with symbolic meaning: ‘a j got stuck in the system.’

2 Email to me, 20 September 2021.

3 The name belongs to my stepfather. He was pictured together with my mother, pointing out family names on the new wall. <https://www.holocaustnamenmonument.nl/nl/home/> and <https://www.volkskrant.nl/nieuws-achtergrond/het-holocaustmonument-in-amsterdam-maakt-het-verleden-tastbaarder~b9d22e10/> (the ‘j’ was eventually corrected, but the ‘ph’ which should have been an ‘f’ was not – I let it go).

4 The stamp was an innovation that the Dutch civil servant Lentz was eager to implement, as part of his broader bureaucratic innovations that helped the Nazi’s operations in The Netherlands. Jurriën Rood, *Lentz. De man achter het Persoonsbewijs*, 2022.

5 Edwin Black, *IBM and the Holocaust: The Strategic Alliance Between Nazi Germany and America’s Most Powerful Corporation* (Crown Publishing Group, 2001).

No ‘j’ was literally stuck in the system this time, of course. My guess is that the explainer was embarrassed and blamed word processing technology. What happened was probably a human mistake we call ‘typo,’ or so the sender’s email address, ‘type o’ (‘tik fout’ in Dutch) suggests. A spelling joke. But typos are still no innocent mistakes to make these days, and incorrect explanations about what happens in ‘systems’ obscures important information about human intent and behavior. We live in times where administrative ‘irregularities’ end up ruining the livelihoods of citizens through inscrutable follow-up computational processing.⁶ Where persons who apply for correction in such cases are not recognized for their accounts of what happened; not thanked for their patience and apologized to, and not served with correction. The newspaper article was not such a case, and the shaky explanation is embedded in an otherwise perfect response: an example of ‘meaningful human review’ as is argued for in current legal, ethical, and societal discourses on what we need to be able to do in times of automated decision making yet which is increasingly hard to accomplish as result of how decision processes are designed. A major focus in these discourses is on explanation challenges posed by increasingly inscrutable, AI-driven knowledge and decision-making tools that support or replace human decision processes.⁷

1.1.2 Explanation as a fundamental condition for humane treatment

Small administrative mistakes are only one processing route by which people end up being treated unjustly. In many cases of what has become known as ‘algorithmic decision making’ (ADM) in public, private and commercial settings, what goes wrong is revealed to be an effect of how the automated decision (support) systems (ADS) are designed to work even if no feeding mistakes are made. As socio-technical organizations, automated systems express and consolidate existing racist and discriminatory hierarchies in our societies, finding new expressions for them. The characterization of such wrongs as accidental is only tenable to some extent, for some groups of people.⁸

6 As will be discussed inscrutability has interrelated human, technological, and organizational causes, all complicating meaningful insight. Whatever the cause, Ranchordás argues that a lack of ‘Administrative Empathy’ in digital times furthermore fuels the victimization of innocent citizens. Sofia Ranchordás, ‘Empathy in the Digital Administrative State’, *University of Groningen Faculty of Law Research Paper* 2021, nr. 13 (2021); See also Scheltema, who argued to augment Dutch Administrative Law’s general hardship clause to exempt citizens from punishment in such cases, inspired by a French law proposal at the time, which afforded citizens ‘the right to make mistakes.’ M. Scheltema, ‘Wetgeving in de responsieve rechtsstaat’, *RegelMaat* 33, nr. 3 (May 2018): 128.

7 The next chapter categorizes, discusses and appraises the challenges that are attributed to the influence and nature of these technologies. Rather than provide readers with a non-representative summary of this discussion here, readers are referred to the next chapter.

8 Patrick Williams et al, ‘Surfacing Systemic (In)Justices: A Community View’ (Systemic Justice,); Agathe Balayn and Seda Gürses, ‘Beyond-Debiasing: Regulating AI and its inequalities’ (EDRi, 2021); Ruha Benjamin, *Race After Technology* (Polity Press, 2019).

The quest, therefore, is for the right kind of understanding of how this happens to enable the right response. Much academic research is done to understand both these things. Depending on the disciplines of the actors involved, their methods of inquiry, the decisional context and various other factors, very different conclusions are drawn and different consequences for the relations of explainers and explainees in decision making processes follow from them. Where some warned early on to focus on understanding and explaining the political embedding of ADS and not let the focus on technological explanations obscure what really matters,⁹ others argued how machine reasoning is more, and at least not less understandable than that of humans and the explanations they afford may therefore “already hit the mark.”¹⁰

In the European legal context that this thesis situates itself in, doing away with human explainers is not an option. In EU and Council of Europe norm setting contexts, the automated handling of the ‘j’s’ in WWII is named explicitly as a reason to anchor a fundamental legal, ethical, and moral duty to provide decision subjects with a human-issued, meaningful explanation of decisions that affect them: a fundamental condition of *humane* treatment.¹¹ Human decision makers (using whatever methods) must avoid to behave like what Arendt described as ‘cogs’: *unthinking* human instruments, or at least uncritical of immoral practices.¹² But it is not yet clear what makes human explanations ‘humane’ with regard to our novel decision making methods. The most cited legal regime for the individual explanation of automated decisions at the time of writing, the General Data Protection Regulation (GDPR) is also the most debated: the relevant provisions fueled so many explanation debates that the GDPR’s ‘right to explanation’ can be referred to as a field of research.¹³ In the meantime, judges,

9 Lilian Edwards and Michael Veale, ‘Slave to the Algorithm? Why a “Right to an Explanation” Is Probably Not the Remedy You Are Looking For’, *Duke Law & Technology Review* 16, nr. 18 (23 May 2017).

10 John Zerilli et al, ‘Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?’, *Philosophy & Technology* 32 (2019): 661–83; Doshi-Velez et al are also optimistic about machine explanations, but draw more tentative conclusions with regard to the future of AI and what needs to be explained about it. Finale Doshi-Velez et al, ‘Accountability of AI Under the Law: The Role of Explanation’ (Berkman Center Research Publication, forthcoming, November 2017).

11 Meg Leta Jones, ‘The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood’, *Social Studies of Science* 47, nr. 2 (1 April 2017): 216–39; The European Group on Ethics in Science and New Technologies (EGE) | EGE - Research and Innovation - European Commission, ‘Statement on Artificial Intelligence, Robotics and “Autonomous” Systems’ (European Commission, March 2018).

12 Hannah Arendt, *Responsibility and Judgment*, Reprint edition (Schocken, 2005).

13 See, among many others, Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation’, *International Data Privacy Law*, 2017; Gianclaudio Malgieri and Giovanni Comandé, ‘Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation’, *International Data Privacy Law* 7, nr. 3 (13 November 2017); Margot E. Kaminski, ‘The Right to Explanation, Explained’, 15 June 2018; Andrew D. Selbst and Julia Powles, ‘Meaningful Information and the Right to Explanation’, *International Data Privacy Law* 7, nr. 4 (1 November 2017): 233–42.

civil servants, doctors, and other explainers grapple with their legal explanation duties while ADS are being implemented in their fields.

1.1.3 What makes a humane explainer?

This thesis is an endeavor to clarify explainers' duties. For a researcher, this entails grappling with different insights and arguments about what defines proper explanation practices, and how such practices' affordances are perceived to be challenged. Over the years, while revelations of algorithmic mishaps accumulated, while governance responses proliferated, some explanation debates started to make more sense than others. Fundamental explanation paradigms under pressure such as those of the Administration, Judiciary, and Medicine were also referred to *as* inspirational sources to flesh out the technologically oriented explanation regimes in place to help them along.¹⁴ But the feed of wrongful ideologies into decision making and its methods existed (long) before and after the establishment of fundamental explanation duties. If explanations have a role to play in preventing such harms from manifesting 'under the radar,' how did we end up here?

A major case of unexplained injustice traveled with the research project, and became a recurring point of reference. In 2021, the Dutch Government resigned over the Childcare Benefits Scandal. By acting mercilessly on the basis of unfounded suspicions of benefits fraud, the Tax Administration¹⁵ had ruined the livelihoods of tens of thousands of parents and irreversibly harmed the childhoods of a multiplied number of children.¹⁶ Over a period of 15 years, parents and childcare providers were tagged as suspects with a mix of manual and digital methods and on the basis of a wide range of attributes. Some were illegal to use (nationality, ethnicity), some consisted of minor 'irregularities' such as small administrative mistakes. The parents had no way

14 'Ongevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen' (Raad van State, 31 August 2018), Kamerstukken II 2017/18, 26643, nr. 557; Marion Oswald, 'Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, nr. 2128 (13 September 2018): 20170359; R.J.N. Schlössels, 'Kroniek beginselen van behoorlijk bestuur 2017', *Nederlands Tijdschrift voor Bestuursrecht* 2017, nr. 9 (2017); AI HLEG, 'Ethics Guidelines for Trustworthy AI', Text (European Commission, 8 April 2019), <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

15 'Tax and Customs Administration' is the English translation that this administrative body uses for itself on its English website. In research and news items about the case, they are also translated as 'Tax Agency', 'Tax Authority', and other varieties. The thesis used their own words minus 'customs', for recognizability: 'customs' is not part of their Dutch name 'Belastingdienst' <https://www.belastingdienst.nl/wps/wcm/connect/en/individuals/individuals>.

16 'Ongekend Onrecht: Verslag van de Parlementaire ondervragingscommissie Kinderopvangtoeslag' (Den Haag: Tweede Kamer der Staten-Generaal, 17 December 2020), 8.

of knowing. It took several parliamentary, commissioned, and civil society inquiries¹⁷ to find out these facts and still more are revealed at the time of writing. And in a very reluctant acknowledgement of what scholars had been pointing out,¹⁸ the responsible State Secretary finally acknowledged that institutional racism was a factor in the Scandal.¹⁹ This fundamental wrong behind the inflicted harms was the last mystery to be cleared up whereas it was arguably the easiest to discover: the disparate impact was obvious. When the acknowledgement came, it was only partial. Rather than discussing how wrongful notions about persons were at play in law and policy, the incidents were flagged as accidents, decision makers as bad apples.²⁰

All those years, all applicable explanation paradigms had failed to prove the truth of what was going on, or any meaningful details.²¹ Over the years of inquiries into what happened, the Scandal discourse revealed institutional confusions about what needs to be justified about what, by whom, to whom, and why that did not happen so that justice could be served. After the Parliamentary Research Commission expressed their special concern that foundational principles had been “reasoned away” by the courts,²² appeal court judges who had presided over cases reflected that they had not done enough to understand the situations, that they should have been alerted by the blatant lack of information in the case files submitted by the Administration, and should have rebelled to the Government and the Administrative Supreme courts about what they recognized as an unfolding constitutional crisis.²³ In other words, they deplored how they had become instruments of an oppressive system. Perhaps, fundamental explanation regulation itself needs more humane points of reference; perhaps the ‘good explainer’

17 Among others, ‘Ongekend Onrecht: Verslag van de Parlementaire ondervragingscommissie Kinderopvangtoeslag’; Besluit tot boeteoplegging Minister van Financiën (Autoriteit Persoonsgegevens 25 November 2021); ‘Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal’ (Amnesty International,), last consulted 26 October 2021; ‘Onderzoek effecten FSV Toeslagen’, rapport (Price Waterhouse Coopers, November 2021).

18 Sinan Çankaya, ‘Opinie | Ze bedoelden het wél zo – het racisme kan onmogelijk ontkend worden’, *NRC*, 27 May 2022, <https://www.nrc.nl/nieuws/2022/05/27/ze-bedoelden-het-wel-zo-het-racisme-kan-onmogelijk-ontkend-worden-a4129407>; Samir Achbab, ‘De Toeslagenaffaire is ontstaan uit institutioneel racisme’, *NRC*, last consulted 1 November 2021, <https://www.nrc.nl/nieuws/2021/05/30/de-toeslagenaffaire-is-ontstaan-uit-institutioneel-racisme-a4045412>.

19 Ministerie van Financiën, ‘Kamerbrief over Fraudesignaleringsvoorziening en vraagstuk institutioneel racisme’ (Ministerie van Algemene Zaken, 30 May 2022), <https://www.rijksoverheid.nl/documenten/kamerstukken/2022/05/30/kamerbrief-reactie-op-verzoeken-over-fraudesignaleringsvoorziening>.

20 And as will be discussed later, the State Secretary stated how institutional racism is not a legal claim, and therefore victims would need to prove they were individually discriminated against. Michiel Bot, ‘Is institutioneel racisme echt racistisch?’ *NJB* 26, 2022

21 Legal explanation paradigms accumulated when parents were effectively forced to enter an array of legal procedures to appeal against various effects of the Tax Administrations decisions on their lives: they were sacked from jobs, their children were placed in custody, possessions were impounded, homes were lost.

22 ‘Ongekend Onrecht: Verslag van de Parlementaire ondervragingscommissie Kinderopvangtoeslag’, 1.

23 ‘Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken’ (Werkgroep reflectie toeslagenaffaire rechtbanken, October 2021).

needs a more explicit profile than that which distancing oneself from past horrors of automation provides them with.

1.1.4 Research objective

The apparent tensions between new problems and old causes inspired to seize the moment for an investigation of what appears to be a weak quality of existing legal explanation paradigms: their propensity to skirt the kind of knowledge making that these times call for so loudly, therewith failing to reveal whether decision making respects decision subjects in their humanity.

The next chapter of the thesis confronts contemporary explanation crisis framings with provocations, and sets out a path of inquiries to sustain explanation regulation moving forward. That chapter therewith further prepares for, and justifies, this thesis's investigation into relations between in/humane qualities of knowledge making as an important dimension of decision practices, how these qualities translate into value-oriented responsibilities for those tasked with explaining decisions, and what this means for the role of legal explanation rules. The chapter builds up to a full set of research questions to match these aims.

1.2 Methodological approach and structure

1.2.1 An iterative research process

1.2.1.1 Between moving targets and sitting ducks

In the starting year of the thesis research the EU General Data Protection Regulation (GDPR) became applicable. Among the many legal and policy challenges that this law came with, 'the right to explanation' was prominent. After a brief debate among legal researchers about whether the right even existed, research on questions around when, why, and especially, *how* Article 22's right would, could, and should

take shape exploded and diversified.²⁴ As moving target #1, tracking this discourse was as interesting as it was challenging. But the extent to which it was interesting changed, or rather, stalled. Several other moving targets, and a gradually materializing sitting duck inspired a course of investigation that departed from the GDPR's specific technological starting point—and from that of other technology focused, emerging regulatory approaches.²⁵

Among the other moving targets was an equally steady stream of research about decisional wrongs produced with ADM. From simple automation to machine learning wrongs that became increasingly harder to identify, case studies of ADM across decisional contexts revealed a growing inadequacy of individually oriented legal protections like that of the GDPR and the human rights regime more broadly—at least in its contemporary iteration. The human rights regime, already tailored to the kind of 'rugged individualism' that underserves people (and indeed communities) in their inevitably relational existence,²⁶ became increasingly more so after the embrace of neoliberalism by those with most influence on fundamental legal developments.²⁷ Especially person and freedom related human rights have come to be interpreted in terms of individual autonomy and dignity, and infringement of them generally requires evidence of individual harm. By abstracting from global human inter-dependency and

24 Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP 251rev.1)' (European Commission, 6 February 2018); Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679 (wp260rev.01)' (European Commission, 11 April 2018); Wachter, Mittelstadt, and Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation'; Malgieri and Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation'; Selbst and Powles, 'Meaningful Information and the Right to Explanation'; Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?'', *IEEE Security & Privacy* 018, nr. 16(3) (2018); Kaminski, 'The Right to Explanation, Explained'; Thomas Hoeren and Maurice Niehoff, 'Artificial Intelligence in Medical Diagnoses and the Right to Explanation', *European Data Protection Law Review* 4, nr. 3 (2018): 308–19; Lee A. Bygrave, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions', SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 29 October 2020), <https://papers.ssrn.com/abstract=3721118>; Emre Bayamlıoğlu, 'The Right to Contest Automated Decisions under the General Data Protection Regulation: Beyond the so-Called "Right to Explanation"', *Regulation & Governance* 16, nr. 4 (2022): 1058–78.

25 The proliferation of 'AI' and other data science governance initiatives in the European regulatory space has become a challenging moving target for researchers more broadly. In rapid succession, the EU alone already issued the General Data Protection Regulation, updated the Medical Devices Regulation to deal with software better, drafted the Data Act, the Data Governance Act, the Digital Markets Act, the Digital Services Act, and the (draft) AI act. Understanding their interplay requires study.. C Codagnone and G Liva, 'Identification and Assessment of Existing and Draft Legislation in the Digital Field', *EU Policy Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies*, 82.

26 Shelley Wright, *International Human Rights, Decolonisation and Globalisation: Becoming Human* (London: Routledge, 2001), 63

27 Samuel Moyn, *Not Enough: Human Rights in an Unequal World* (Cambridge, Massachusetts: Harvard University Press, 2019); Those most knowledgeable of the harms that ensue are frequently with the least resources to engage in legal battles, with which their experiences don't inform legal developments. Williams et al, 'Surfacing Systemic (In)Justices: A Community View'.

structural oppression,²⁸ the propensity of ADS to express and consolidate racist and discriminatory societal hierarchies is inadequately addressed.²⁹ Along the way, an increasingly sophisticated, multi-disciplinary record of decisional processes, aspects, and outcomes that were seen to require justification established.³⁰ Technological aspects were among them.

In discourse around what aspects require what kind of justification (ethical, legal, moral), and what ‘technological’ explainability that requires, the technological field itself produced much research, too: moving target #3. Conferences like FAccT (Fairness, Accountability, and Transparency, formerly ‘FAT’)³¹ convened many of these fields under one umbrella and became an important source among more singularly oriented conferences (which themselves started to diversify however, which also required attention.)³²

In this research, and in supra-national governance documents that were being produced, fundamental Administrative principles of ‘motivation,’ ‘due diligence’ and ‘due process’ were much named as inspiration of what justifying decisions means and what needs to be explained about what to do so. Another much named domain was medicine, both for its ethical principles and for its informed consent paradigm. While engaging with the above-named bodies of research, I had indeed started to look at Dutch Administrative Law’s explanation rules for ‘inspiration’ of what this salient explanation domain would require qua *technological* explainability, and what that could teach us for other fields. The second domain (medicine) was scheduled to follow.

After years of working in the legal aid field, my positive expectations with regard to the usefulness of the first explanation paradigm were however not very high. In terms of inscrutability some manual administrative decision making did not under perform in comparison with complex ADM. But to what extent that was down to codified rules, policy, compliance issues, or a lack of legal process protections remained to be seen. Especially the role of explanation rules was unclear to me. In other words, I was ‘biased,’ but I also stepped in with open-minded curiosity.

28 C. McCrudden, ‘Human Dignity and Judicial Interpretation of Human Rights’, *European Journal of International Law* 19, nr. 4 (1 September 2008): 655–724; Wrongful conceptualizations of autonomy and dignity have also supported wrongful directions of medicine, see for example Charles Foster, *Human Dignity in Bioethics and Law* (Hart Publishing, 2011).

29 Anna Lauren Hoffmann, ‘Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse’, *Information, Communication & Society* 22, nr. 7 (7 June 2019): 900–915.

30 Maranke Wieringa, ‘What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability’, in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* ’20 (Barcelona, Spain: Association for Computing Machinery, 2020), 1–18.

31 <https://facctconference.org/>.

32 For example, the yearly Computers, Privacy and Data Protection conference (CPDP).

This first exploration gradually produced the image of the sitting duck. A relatively brief exploration of the much-lauded rules & principles revealed them to be inadequately framed to grapple with the *kind* of harms that ADM was seen to produce—also when ADM was not a factor. And since ADM’s harms do not affect people equally but swells relations informed by racism, discrimination and marginalization, the need for a ‘deep dive’ emerged. If these, and possibly other or all, fundamental explanation paradigms were badly designed, ‘fixing’ them with additional technology-dependent regulation would help.. who? Put differently, whose problems were apparently not seen to be in need of ‘fixing’ before?³³ The domain findings were parked to be analyzed later.

The choice to focus on explainers’ obligations, rather than explainee’s needs, was made in this period. In legal ADM explanation developments, human ‘reviewers,’ decision maker-cum-explainers, were situated as guardians of fundamentally humane treatment. The human rights based, moral load of their roles was emphasized,³⁴ with which an appeal to reviewers’ supra-legal conscience was expressed: to avoid to become instruments of unjust law and policy. The choice to focus on their ‘fundamental’ obligations therewith fits with a view of law as ‘ethical’: of legal order as a “complex of human activities” which brings in a “final responsibility” to ethics and morality.³⁵ Put differently, with an understanding of the ‘essence’ of law as fundamental obligations that come with ‘being a person.’³⁶ The choice also fits with an understanding of law as an instrument that aims to establish ‘mutual trust’ between all persons and therefore needs to include all their interests, deserve (and attract) all persons’ respect.³⁷ The choice to act on the ‘suspicion’ that explainers are not instructed very well also fits with an understanding of how law repeatedly failed to live up to these ideals. Taking the fundamental doubts about the beneficence and force of fundamental legal explanation rules and principles seriously, most earnestly phrased, requires a “refusal—as a matter of research integrity—to conduct such research in the service of aspirations to wield power over others.”³⁸

33 Put differently, perhaps ‘[i]t’s not that law is failing to regulate the harmful effects of algorithms, but rather that algorithms are exposing the comprehensive failure of the law to address real in justice.’ Dan McQuillan, *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence* (Bristol University Press, 2022), 39.

34 Jones, ‘The Right to a Human in the Loop’.

35 Felix Cohen, ‘The Ethical Basis of Legal Criticism’, *The Yale Law Journal* 41, nr. 2 (1931): 205; Calo writes about how this is also what ‘law can share back’ with science and technology studies, or STS: ‘law’s ‘relative comfort with normativity, second to none in academia only to moral philosophy.’ Ryan Calo, ‘The Scale and the Reactor’, SSRN Scholarly Paper (Rochester, NY, 9 April 2022), 17, <https://doi.org/10.2139/ssrn.4079851>.

36 Ernst Hirsch Ballin, *Advanced Introduction to Legal Research Methods* (Edward Elgar Publishing, 2020), 70.

37 Ernst Hirsch Ballin, *Advanced Introduction to Legal Research Methods* (Edward Elgar Publishing, 2020), 70.

38 Hirsch Ballin, 94; See also Cohen, ‘The Ethical Basis of Legal Criticism’, 205.

These choices were sustained by a gradually deepening engagement with a (mainly) philosophical body of work concerned with how relations of knowledge and power express in human treatment and what that means for the definition of acting ‘justly.’ The fields of epistemic injustice and justice are well poised to address the kind of harms that ADM is seen to produce. This research informed my ‘seeing’ and therewith my angle of investigation.³⁹ It was eventually applied into a critical research tool to sustain the more fundamental decision domain research that needed to be done: to qualify the findings of the first domain, and to inform the research on the second domain.

But my recognition of the usefulness of this body of work, and choices with regard to what to engage with it for (and how) was also produced *by* the other lines of investigation that I had set out. In other words, the research process of this thesis was far from linear. It entailed a going to and fro between historical sources (on the harmful effects of automation, on explanation rules), contemporary sources (domain explanation frameworks and the moving targets), and epistemic in/justice literature that were themselves interestingly developing and whose understanding took time.

How to report on that? The choice was made to dedicate a chapter to the all the aspects named above. A chapter that justifies the thesis’s choices and focuses, embedded in *a kind of* ‘literature review,’ or a ‘review plus’: a discussion of what it means to ‘source research questions from a tension,’ anchored in sources from early and later stages of research, from all cited bodies of work; showing how engaging with these produced the eventual research questions.

1.2.1.2 *Disconnecting from automation*

There is one other ‘moving target’ that needs to be named at this point. The introduction already introduced the Childcare Benefits Scandal: a Dutch Administrative-Judicial-Political miscarriage that gravely and irreparably harmed many thousands of parents and children. By sheer accident I have been an early witness: cases had started to trickle in at the legal aid office where I worked at the time, challenging the understanding of the lawyers there. Much (if not all) of what actually took place in the Tax Administration’s decision processes remained unclear throughout the many years the scandal played out for. The same was true for the apparent failing of the justice system. But this changed during the thesis’s research years. An amalgam of causes surfaced: discriminatory wrongs in manual and automated steps in group and individual decision making; too much and too little political steering; and a judiciary that gradually landed

39 Peter Rule and Vaughn Mitchell John, “A Necessary Dialogue: Theory in Case Study Research,” *International Journal of Qualitative Methods* 14, no. 4 (November 20, 2015): 1609406915611575

in an identity crisis.⁴⁰ These revelations only established over (many) years, after institutional and civil society investigations, and through discussions that followed after conflicting reports. As was described in the introduction, a core aspect of what happened was only acknowledged, and only partially, at the very end of my trajectory: the institutional character of the racism and discrimination at play.

All systems of justification had failed to prove the truth of this. In the course of the years, the ‘Scandal discourse’ revealed institutional confusions about what needs to be justified, about what, by whom, and why that did not happen so that justice could be served. This discourse is sourced as a red thread throughout the thesis. Some chapters cite judges’ reflections, others cite discussions of (disagreeing) legal scholars in a reflection on the functioning of Dutch Administrative Law rules. Yet others reflect on how the scandal was the biggest, but certainly not the first in its kind in the Netherlands, sustaining the thesis’s premise that naively calling wrongs out as ‘novel,’ and pointing the finger to technology, is not a way forward. The scandal discourse therewith strengthened the choice to ‘disconnect from automation’; to abstract from automated decision making in the domain analyses that were eventually performed. To cite Lorraine Code, whose work became an important source of the thesis’s theoretical framework: “[i]n our quests for understanding, the appropriate questions broaden from “what can we know” to “what sort of discourse does the situation really call for?”⁴¹

1.2.2 A critical analysis of legal explanation rules

1.2.2.1 Outwith law

The type of investigation of legal explanation rules that the thesis was seen to require is critical, fundamental, and to some extent exploratory. Critical in the sense that existing rules need to be placed under scrutiny on the suspicion that they fail to serve particular values, with which they are an obstacle towards progress.⁴² Fundamental, since that is what these values are. That means that an ethical, moral position outside law needed to be designed; a position that allows to qualify law. At the same time, the description of the legal explanation paradigms that needed to be assessed required to step inside of law, too: to explain *its* explanation aims in light of a domain’s particular decision making. That part is usually approached with ‘doctrinal methods’: hermeneutical descriptions of law on the basis of contemporary and historical legal sources (rules, principles, case-law, parliamentary history..). After some more words on the need for an angle from outside, the remaining sections discuss the chosen angle, the choice of

40 ‘Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken’. Reflectierapport van de Afdeling Bestuursrechtspraak van de Raad van State’, overzichtspagina (Afdeling Bestuursrechtspraak van de Raad van State).

41 Lorraine Code, *Epistemic Responsibility* (Brown University Press, 1987).

42 Hirsch Ballin, 30.

domains, the ‘doctrinal light’ approach of the descriptive parts of the domain chapters, and what that meant for the approach of each domain.

The choice to wrestle free from established legal uses and understandings of fundamental values and concepts such as dignity, autonomy, and from typical legal understandings of justice (e.g. distributive, procedural, restorative) was done to attempt to wrestle free from the possibly inadequate or even suspect political load of them. This region’s legal rules, and the principles that rule the rules, have roots in traditions that served expansive and oppressive colonial political powers.⁴³ The thesis seeks to know whether and how the kind of values and interests that follow from a radically solidary notion of ‘explanation justice’⁴⁴ are ignored in them. It also wants to find out how such norms *could* and *should* be acknowledged, and what general and domain-specific grounds for this can be identified.⁴⁵ This ambition is co-inspired by Benjamin, who argues a necessary re-imagining of science and technology in order to liberate it from racial oppression.⁴⁶ Among other creative approaches to this, she references legal scholar Bell’s literary methods (such as ‘reversals’ of white and black roles) who argued that “to see things as they really are, you must *imagine* them for what they might be.”⁴⁷ These re-imaginings of critical race studies remain urgently necessary, and it is with humbleness that this thesis hopes to contribute to a ‘re-imagining’ of explanation rules to serve this and other critical angles.

Inspiration for the chosen fundamental-critical-exploratory approach, and for the approach of the domain studies was additionally found in a related discipline of policy analysis; more precisely in Bacchi’s methodological approach named ‘What’s the problem represented to be?’⁴⁸ The purpose is to study ‘problematizations,’ problems that are addressed in policies but that are, more importantly, also *created* in policy. Policy is studied here as a politically informed knowledge-making practice,⁴⁹ which

43 Wright, *International Human Rights, Decolonisation and Globalisation*; As Martin Conway discussed recently in Amsterdam, historians have tended to keep with a narrative of the shaping of constitutional democracy after WWII that ignores how the colonial horrors before, and those of decolonization after, have been deliberately kept out of political discourse (& education) by lawmakers and other shapers of post WWII European Democracy. Martin Conway, ‘The Certainties of the Past? The Making of Democracy in Western Europe after 1945’ (Amsterdam, 26 January 2023), <https://spui25.nl/programma/the-certainties-of-the-past>.

44 ‘The basic concepts of law, including those of the constitution, must be retraceable to the founding notion that every human person deserves recognition.’ Hirsch Ballin, *Advanced Introduction to Legal Research Methods*, 94.

45 Hirsch Ballin, 24, 87.

46 Benjamin, *Race After Technology*, 195.

47 Benjamin, 195.

48 Carol Bacchi, ‘Why Study Problematizations? Making Politics Visible’, *Open Journal of Political Science* 2, nr. 1 (26 April 2012): 1–8.

49 Bacchi uses ‘practice’ in a Foucauldian sense: as places, where “what is said and what is done, rules imposed and reasons given, the planned and the taken for granted meet and interconnect.” Practices are “intelligible backgrounds” for the kind of thought, truths, and politics they produce. Bacchi.

it is as well. Just like law. ‘Explanation’ emerges as a legal problem through how law defines a need for it.

As a research tool, problematizations serve different research traditions. Among others Bacchi cites Paulo Freire’s development as “a strategy for developing a critical consciousness” by pulling accepted “truths” into question as “myths fed to the people by the oppressors.” Revealing how such truths are made allows to meaningfully engage with them. Bacchi modeled her method on that of Foucault, who developed his to reveal the socially powered relations that produce our world as we ‘know’ it and, and as it ‘knows’ us.⁵⁰ Similar to Foucault’s analysis of ‘practice artifacts,’ Bacchi’s method presents policy proposals and other documents as texts that are *prescriptive* for the subjects/objects of that policy: governing takes place *through* particular problematizations,⁵¹ and therewith express a society’s ‘govern-mentality.’ The study of problematizations, she argues, allows for innovative research strategies that bring “complex strategic relations that shape lives,”⁵² meaning politics, into view.

1.2.2.2 The critical angle of choice: epistemic injustice and justice literature

Selecting theory to inform the critical angle with which to perform these tasks was itself a critical task. There is a need to understand to what extent such theory (already) needs to speak to the subject, and there are integrity related questions around the required depth of engagement, respect for original context, requirements for applicability in a different context, and to what extent this application contributes to the original theory’s further development.⁵³

The search was for a broader theory or set of theories that would sustain and respect the more particular critical angles that are practiced such as critical race, feminist, disability, anti-colonial theory and allows to cite relevant authors while avoiding to wrongly appropriate their findings.⁵⁴ The ‘reading journey’ started with a focus on positive obligations: what it could mean to do right by explainees. A first selection of works on epistemic justice was sourced. These describe proper, ‘virtuous’ knowledge practices in acknowledgment of the inherent politics of knowledge. Justification is a

50 Bacchi, 1,3.

51 Bacchi, 5.

52 Bacchi, 1.

53 Peter Rule and Vaughn Mitchell John, ‘A Necessary Dialogue: Theory in Case Study Research’, *International Journal of Qualitative Methods* 14, nr. 4 (20 November 2015).

54 A text of Mitova, which I found on my path much later importantly helped to cross-validate the journey that led from particular critical angles to epistemic injustice theory. Veli Mitova, ‘Decolonising Knowledge Here and Now’, *Philosophical Papers* 49, nr. 2 (3 May 2020).

major theme. The literature describes norms for ‘technical’ methodological dimensions alongside behavioral norms for humans engaged in knowledge making processes.⁵⁵

Increasingly, the inverse of this became the focus: what it could mean to do wrong by explainees. Understanding this allows to ‘get the most’ out of the practical guidance derived from the positive norms in justice-oriented literature, but perhaps more importantly, serves to corroborate whether the right values and norms were identified in the first place: whether these are indeed norms that *aren’t* followed when epistemic authority is misused. Translated to the thesis context: when explanatory relationships ‘malfunction.’

The justice and injustice domains are obviously related, and cross-cited by authors in them. But not always, and not all literature cited in Chapter 3 is categorized as epistemic justice or injustice to begin with. The understanding of the consulted fields, how they do and do not hang together, and when a work can be categorized as belonging to either was built up along the way. The selection of texts was made dependent on authors’ explicit engagement with fundamental equality, interdependence, fairness, dignity, and justice with regard to ‘how we know’ or in philosophical terms, epistemology. The difference with the broader field of ‘social epistemology’ lies in how that field does not *necessarily* commit to this. To compare with the STS fields: where these always investigate the sociality of technological practices, they do not necessarily seek to understand its politics or morality.⁵⁶ And as various cited authors argue, the other fields of social epistemology are still closer to the highly non-diverse and euro-centrist origins that apply to philosophy more broadly.⁵⁷ The geopolitical dimensions of epistemic justice and injustice literature itself is a subject of interest and study—and needs to be that in light of the dominance of ‘North-Western’⁵⁸ thought in Academia itself, and the Colonial roots of its knowledge practices.⁵⁹ Through engaging with *injustice* literature the geographical scope of the investigation broadened decisively. This led to the deeper understanding of the fields as described above, and to a more responsible use of the earlier found epistemic justice materials. *The Handbook*

55 Machteld Geuskens, ‘Epistemic Justice: A Principled Approach to Knowledge Generation and Distribution’ (Tilburg University, 2018); Bernard Williams, *Truth and Truthfulness* (Princeton University Press, 2002); Code, *Epistemic Responsibility*.

56 And therewith do not necessarily or easily sustain legal scholarship. As Calo wrote, ‘STS is beautiful music, but can you dance to it?’ Calo, ‘The Scale and the Reactor’.

57 Charles W. Mills, ‘White Ignorance’, in *Agnology: The Making and Unmaking of Ignorance*, edited by Robert N. Proctor en Londa Schiebinger (Stanford University Press, 2008), 230–32; Heleen Booy en Kiki Varenkamp, ‘Diversiteit als toetje: het filosofie-examen gaat over 44 mannen, vier vrouwen en één filosoof van kleur’, *De Groene Amsterdammer*, 2021.

58 As Tuck and Yang argue, the phrase ‘North Western’ itself is misleading as it excludes indigenous knowledges in the North-Western hemisphere. Eve Tuck and K. Wayne Yang, ‘Decolonization Is Not a Metaphor’, *Decolonization: Indigeneity, Education & Society* 1, nr. 1 (2012).

59 Chen Bar-Itzhak et al, ‘In Search of Epistemic Justice: A Tentative Cartography’, *University of Pennsylvania international CFP listing* (blog), last consulted 13 December 2021, <https://call-for-papers.sas.upenn.edu/cfp/2021/12/09/in-search-of-epistemic-justice-a-tentative-cartography>; Priyamvada Gopal, ‘On Decolonisation and the University’, *Textual Practice* 35, nr. 6 (3 June 2021): 873–99.

of *Epistemic Injustice* was an important source, also in terms of how it showcased possible approaches of relevant themes such as “authority, credibility, justice, power, trust, and testimony.”⁶⁰

1.2.2.3 Theoretical application

In a reflection on her 1987 *Epistemic Responsibility*, Code recalls how some scholars at the time found that she should have provided more ‘ready-made’ solutions to the challenges of proper knowledge making that she had described. But this had not been her aim, at least not in the form of ‘necessary and sufficient conditions.’ She had built “impressionistic guidelines.”⁶¹ For her thesis, this was arguably the more responsible choice. It expresses her argument that knowledge practices should not settle too easily on what counts as necessary and sufficient at any time.

This thesis needs to make a more practicable contribution. That is not to say that criticizing law is not a valuable normative enterprise in itself. Indeed, one value of this thesis’s theoretical development is to offer a ‘thinking tool.’⁶² The thesis hopefully allows to imagine what an explanation practice that aims for knowledge justice could look like. But as was explained in the introduction to this chapter, the eventual aim is to inform explanation regulation in AI-infused times. For this, guidance needs to be modeled more explicitly.

To that end, Chapter 3 translates and applies the consulted theories into a model comprised of ‘duties of care’ (hereafter: modeled duties of explanation care, or Model). It describes obligations for explainers as they move through four phases of a theoretical explanation cycle: investigation of their own information positions; of those of their explainees; interaction with their explainees, and record taking. The description of these duties themselves go beyond ‘impressionistic,’ but they are still guidelines. They are meant to be further developed, further ‘imagined’, and translated to context-sensitive instructions for explanation domains. To that end, the tool is ‘ready-made’ and usable to perform explanation domain research with. An analysis on the basis of the Model allows to assess the epistemic justice aims, and epistemic justice potential of existing legal explanation rules. Its function as a thinking tool is a little bit different here: it should be installed in a researchers mind while they are sourcing materials with which to describe a domain’s decision-making authority, and explanation regulation. The questions it inspires to ask are about a practice’s ‘problematization.’ What decisional powers are attributed here? Why do people supposedly need an explanation? Why,

60 Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., ‘Introduction to The Routledge Handbook on Epistemic Injustice’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017), 1.

61 Critics were left wanting for “a set of accompanying rules for the direction of the mind“ Code, *Epistemic Responsibility* (2017) In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.

62 Rule and John, ‘A Necessary Dialogue’.

and how, was such a duty regulated? How is it expressed in language? What ‘counts’ as a proper explanation to serve these ambitions? The last question is not really one question. As Bacchi advises, problematizations tend to multiply. Further, ideally empirically informed explorations could look at how, within a practice, explanation rules work out.⁶³ When are they seen to be sub-optimally implemented, and why? The point at which to stop asking, as always, is a question in itself. It is also one of scope. For this thesis, at least the first five questions should be sufficiently answered.

1.2.3 Approach of two regulated explanation domains

1.2.3.1 Selection and focus

The decisional practices whose explanation rules are critically analyzed are set within Administrative Law and Health Care. Within these domains the scope was further delineated to, and/or the spotlight directed on, situations of large(r) power and information imbalances in combination with social dependency. This translates to e.g. attention for welfare decisions in the Administrative Domain, and within Medicine the scope was narrowed to General Practice: Dutch citizens’ traditional first stop for health complaints and a necessary stop for specialist referrals. Further in-domain choices and considerations are discussed in the domain chapters, themselves.

The domains were selected firstly for their fit with two, cumulative conditions. Firstly, the relation of decision makers and explainees is a commonly experienced one. These are well-known rules, practiced widely on a daily basis and therewith a substantive societal expression of our ‘govern-mentality.’ Secondly, it is an explanation relation in which the information inequality between the two parties is large and matters non-trivially because of the social dependency of explainees on the outcome of the decisional process for their well-being and thriving. Explainees’ well-informed participation in these decision practices is of utmost importance.

The selected domains are also ‘foundational’ in the sense that they are both much named as fundamental explanation paradigms. They are exemplary, and named as such in ADM explanation literature. Both domains also have long standing engagements with automation and technological innovation, and are facing explanation challenges

63 Hirsch Ballin, *Advanced Introduction to Legal Research Methods*, 42.

related to the advances of AI.⁶⁴ All the ‘explanation challenges’ that are discussed in Chapter 2 are at play in them. Importantly, they *have* been at play in them, independent of automation’s challenges. Lastly, the two domains together provide a ‘broad spectrum scan.’ The Administrative domain is a prototypical rule based decision-domain, and Medicine (or Health Care) is a prototypical expertise-based domain. The role of (explanation) law in the domains is different because of this, which allows to describe different kinds of dynamics.

Notwithstanding the focus on the reviewer in many ADM explanation governance instruments, this thesis’s study of legal rules focuses on the obligations of ‘first explainers.’ The point is to understand how their instructions express values that the thesis accepts as ‘basic’ as well as ‘fundamental’: they need to be ‘always on,’ not added as a bonus for decision subjects that are able to file for review or judicial appeal. As the next section will explain, this has consequences for *how* the rules were studied.

64 Robin Pierce, Sigrid Sterckx, and Wim Van Biesen, ‘A Riddle, Wrapped in a Mystery, inside an Enigma: How Semantic Black Boxes and Opaque Artificial Intelligence Confuse Medical Decision-Making’, *Bioethics* 36, nr. 2 (2022): 113–20; Ryan Calo and Danielle K Citron, ‘The Automated Administrative State: A Crisis of Legitimacy’, *Emory Law Journal* 70, nr. 4 (2021); Not just for doctors themselves. The ever-increasing powers of tech companies also brings in different research ethics. For example, several medical researchers call out tech companies for how they present inscrutable research results on purpose ‘AI Is Wrestling with a Replication Crisis’, *MIT Technology Review*, last consulted 13 November 2020, <https://www.technologyreview.com/2020/11/12/1011944/artificial-intelligence-replication-crisis-science-big-tech-google-deepmind-facebook-openai/>.

1.2.3.2 *Decisional whats, who's, hows, and explanation rules: descriptive parts of the domain chapters*

For the descriptive parts of the domain research chapters, traditions of doctrinal legal research⁶⁵ were followed ‘lightly.’ Rather than a full-on doctrinal mapping, a more functional description was pursued.

The first part of each chapter characterizes the decision making at hand. These parts are structured in ‘what’, ‘who’ and ‘how’ sections. These parts pay attention to how the decision making is (also) political; describe the decisional practices in terms of how they are (also) knowledge making practices, and characterize the domains’ decision makers in terms of what is expected of them in light of these two dimensions. Choices with regard to consulted literature are explained in the respective chapters.

The second part describes each domain’s main explanation rules. This is where the ‘doctrinal light’ characterization matters most. The point of the chapters is not to provide a comprehensive legal mapping that e.g., places the rules in their international context, discusses historical and contemporary case-law developments, and takes such judicial interpretations ‘as a lead.’ The discussion rather focuses (as stated earlier) on the *direct* guidance of the rules for *initial* decision makers: how they are instructed by them in rules that are publicly known (and *supposedly* known by explainees), how this relates to what was described about the knowledge-and-decision making that these explainers are authorized to do, and the (type of) moral/ethical considerations this was described to come with. Codification history, legal developments and (multi-disciplinary) scholarly critique are engaged with to describe the explanation rules’ aims, again in relation to what was described about what happens in the fields. In that sense, the explanation rules are certainly discussed in terms of ‘proportionality to their purposes,’⁶⁶ but the main purpose of the chapters is on the rules’ proportionality to the ‘re-idealized’ purposes that are embodied in the modeled duties of explanation care that the thesis developed.

Originally, the desk research of both domains was to be complemented by a qualitative investigation in the form of expert interviews with explainers. This plan had to be abandoned after the Coronavirus pandemic hit the Netherlands. This happened a few weeks into the research for the medical domain. Persons, research locations, and even materials became scarce as hospitals (where many materials were kept) were inaccessible, and medical specialists’ schedules became overburdened. A series of four

65 “synthesis of various rules, principles, norms, interpretive guidelines and values .. [which] explains, makes coherent or justifies a segment of the law as part of a larger system of law.” Terry Hutchinson and Nigel Duncan, ‘Defining and Describing What We Do: Doctrinal Legal Research’, *Deakin Law Review* 17, nr. 1 (1 October 2012): 117.

66 Hirsch Ballin, *Advanced Introduction to Legal Research Methods*, 89.

informal, preparatory interviews were still conducted.⁶⁷ When it became clear how long the situation would last, the empirical plans were dropped.

1.2.3.3 Qualification of the explanation rules in terms of the thesis's theoretical development

The third part of each domain chapter engages with the preceding two parts and draws them together by analyzing the explanation rules in terms of their 'justness' potential: an analysis guided by the 'modeled duties of explanation care' that were developed in Chapter 3. Each of the Model's four elements is treated in a separate section. Each section first discusses the pertinence of, and recognition of, the elements' aims and values in each field of decision making. This part considers if and how the Model's aims 'make sense' in, and for, the field, and how they were engaged with in literature about it. These findings are then related to the expression of the aims and values (or lack thereof) in the domain's explanation rules: part two of each section. With this, a critical description of the domain's legal explanation paradigm is constructed. The last chapter of the thesis draws lessons from these studies to inform the (further) development of ruled explanation paradigms in AI-informed times.

1.2.4 Contribution of the thesis

The contribution of this work lies on several planes. Firstly, the thesis reveals how fundamental legal paradigms and principles that are appealed to as the ultimate benchmark with regard to humane, dignified treatment of explainees are not 'fit for purpose.' With this the thesis hopefully helps to prevent that above ground solutions to AI-infused explanation challenges underperform or, worse, obscure these fundamental flaws.

Secondly, the thesis engages with philosophical fields that are increasingly being recognized as highly relevant, informative and useful for doing work on justice-related aspects of data technologies and artificial intelligence. This recognition is however mostly found in non-legal discourse, and clear examples of how the philosophical and legal disciplines can be usefully related on this subject are hard to find. The thesis contributes to the construction of such bridges.

Thirdly, the thesis makes a practical contribution with which further theoretical as well as empirical research across decisional domains can be directly sustained. By building a normative framework and using it as a research lens to formulate general conceptual criticism as well as making more practical, detailed points, the thesis demonstrates what the work of improving regulated explanation paradigms could look like.

67 I am especially thankful to Bob de Groot for sharing his historical knowledge about explanation duties in GP practice and professional norms, to Marco Philipoom for insight into the contemporary education of GPs, and Anne-Sara Breur for sharing her learning materials and experiences about informed consent.

Lastly, and perhaps most fundamentally, the thesis hopes to strengthen the positions of those tasked with justification and explanation. These persons are put forward as ultimate guardians against oppressive decisional practices, but they are not sufficiently equipped for this role. This does not just pose a risk for individual decision subjects but invites precisely those grand-scale wrongs that explanation regulation says it means to avoid.

1.2.5 Thesis structure

Chapter 2 positions the thesis project in relation with contemporary, multi-disciplinary research on challenges to fundamental values that are perceived to be posed by increasingly inscrutable technological decision (support) methods. After a brief introduction of (general and fundamental) legal aims of explanation regulation, it categorizes the perceived challenges, and proceeds to discuss how these challenges are real but not as new as they are made out to be. The preexisting challenges have mainly affected decision subjects from non-privileged groups. An argument is built for the need to investigate existing explanation regulation in terms of (lack of) progress, and conclusions are drawn about what kind of investigation this calls for.

Chapter 3 engages with literature from the philosophical fields of Epistemic Justice and Injustice, using these to perform a ‘re-idealization’ of explanation obligations. The chapter discusses insights about how knowledge practices can be conducive to social oppression, and what preventative and restorative labor needs to be engaged with in response to prevent and fight this. It then translates these insights to the relation of explainers and explainees, and ends by modeling the insights into a set of four key ‘duties of care’ for explainers in institutionalized explanation practices.

Chapter 4 and chapter 5 test the modeled duties of explanation care on the main legal explanation regimes of two decision domains of different character: rule based Administrative Law and expertise based General Medical Practice. It first describes these domains in terms of what decisions are made there, what characterizes decision makers, and what are the relevant, significant methodologies. In these descriptions automation is mostly abstracted from, in line with the chosen investigative focus on how explanation regulation was set up to deal with similar but ‘analog’ versions of the identified problems. The middle parts of the chapters discuss each domain’s main legal explanation rules, showing the laws’ engagement with what was described about the what, who, and how of each domain. Part three of each chapter qualifies this engagement in terms of the modeled duties of explanation care.

Chapter 6 draws lessons from both domain studies, arguing that these should inform the (further) development of ruled explanation paradigms in AI-informed times. It discusses relevant ways in which ‘islands of recognition’ for the modeled values and objectives that were found in research about both domains are actively or passively

frustrated in the domains' legal explanation rules. It relates these observations to ongoing ADM explanation regulation efforts, explicating how the thesis's analysis can support the work of explainers, researchers, and rule makers in AI-infused times: times in which shying away from discussing knowledge in explanation is rightly, but belatedly, problematized. The chapter is structured in four sections that are each dedicated to one of the Model's four elements. Takeaways in the form of observations are discussed for each domain consecutively, preceded by general observations.

Care to explain?

2 New crisis, old problems: sourcing research questions from a tension

This chapter describes and delineates the problem space, further introduces the thesis's focus and subject, justifies and clarifies the thesis questions, and introduces a (mainly) philosophical field of research that was engaged with throughout the research project. Section 2.1 introduces the phenomenon of regulated explanation, and the types of situations the thesis focuses on. Sections 2.2 and 2.3 discuss how this paradigm is seen to be challenged, and why the presentation of these challenges as new deserves to be questioned, respectively. Section 2.4 draws conclusions with regard to what that should mean for the focus and directions of such questioning. It engages with uses of 'meaningful' in regulatory solutions to the perceived explanation problems, then formulates its own two notions of 'meaningful' to pursue. It settles on the explainer as actor to focus on, and paves the way for an investigation into explanation rules as instructive (for who uses them), expressive (of societal values), and prescriptive (with regard to how decisions can be made.)

2.1 Explanation rules and aims: a very short introduction

2.1.1 Regulated explanation in situations of substantive power and information inequality

What follows is a very brief introduction to the phenomenon of legal explanation rules, and the type of ruled situations that the thesis focuses on: institutionalized decision practices with considerable power and information inequality between decision makers/explainers and their explainees. The point of the section is to create just sufficient familiarity with the subject to usefully understand the critical discussion of contemporary explanation concerns in this chapter. More in-depth understanding of established, and possible, rationales and ideals for regulated explanation is established throughout the thesis.

With 'explanation rules' this thesis refers to rules that determine what needs to be explained to individuals about decisions that affect them. Such rules exist in law, and in other governance instruments such as professional ethics, codes of conduct. Describing explanation rights and corresponding duties, such rules are designed to fit different decisional domains and the kind of explanation that is aimed for in them. The main focus of the thesis is on those in law. The types of decisions that by law need to be explained are not all colloquially referred to as decisions, however. Think of medical treatment recommendations that patients need to be informed about in order to consent. This example also illustrates how the term explanation, itself, is not always used in an explanation rule. A judicial sentence is typically called a motivation, and explained as a justification. As will

be discussed, even settled definitions of decisions (such as a welfare eligibility decision) are being *unsettled* because automation adds decisional steps that unsettle what to label as decisive moments that explanation rules should see to.⁶⁸

The thesis creates unity by stubbornly speaking of ‘decisions’ and ‘explanations’ in the regulated situations it focuses on. Simply put (as this will be discussed in detail later), it treats decisions as conclusions, and understands explanations as reasoned accounts of such conclusions. Depending on the breadth and depth of a domain’s explanation rules, these accounts cover the why’s, and or how’s, of decisions. This allows decision subjects and others to assess whether the (proposed) treatment of a subject agrees with a society’s rules and other norms. Explanation *rules*, with that, are expressions of what a society considers to be of interest to know about a decisional process, and what counts as sufficiently reasoned.

The three examples named (Judiciary, Administration, and Medicine) are types of domains the thesis is especially concerned with. Both power and information inequalities are typically large in these domains. Insofar as needed the thesis furthermore prioritizes situations in which decision subjects are especially in need of support and so especially vulnerable to unjust treatment. This is somewhat of a given in the medical domain, but for the study of the Administrative domain this means the focus will be on sub domains such as welfare eligibility rather than, say, permissions to organize an event. All this is not to say that what are typically regulated as more equal, or ‘horizontal’ relations are not also loci of covert or unexplained power abuse; indeed the ‘platform’ economy’s algorithms have made them more so.⁶⁹ The argument this thesis builds for explanation care will be applicable there, too, as it is not domain dependent. The reason that the thesis focuses on these other situations is because in these domains, explanation is already seen to perform a fundamentally important function against oppression, and is referred to as benchmarks in discussions of ADM challenges.

Broadly speaking, explanation regulation in these domains is focused on the justification of decisional powers and ‘power of expertise,’ that is on ameliorating the information imbalance that also exists between decision maker and decision subject. As was mentioned, explanation rules express as rights of decision subjects and duties of decision makers. Rules don’t necessarily distinguish between the two in how they are worded; a rule can for example state that ‘upon request, the decision subject is provided with information.’ This thesis studies explanation rules as duties in the sense that it concerned with how they instruct explainers, either directly or implicitly, to serve the needs of their explainees. Implicitly for example via the description of explainee’s rights, but also through how explanation rules may simply oblige that a decision ‘is explained,’ addressing who is accountable for making the decisions. This person does not always conflate with the

68 Reuben Binns and Michael Veale, ‘Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR’, *International Data Privacy Law* 11, nr. 4 (2021).

69 Karen Levy and Solon Barocas, ‘Designing Against Discrimination in Online Markets’, *Berkeley Technology Law Journal* 32 (2017): 1183–1238.

explainer (or even with the actual decision maker). A medical specialist's finding may be communicated via their assistant or a GP, or an administrative body decision may have been made 'fully automatically.'

2.1.2 Clarification, explanation, justification: instrumental aims of explanation rules

This section further introduces some common instrumental aims of explanation rules. Instrumental here means that explanations are aimed to achieve a particular result, as opposed to 'explaining per se' as a matter of respect for the humanity of decision subjects – an intrinsic value of explanation.⁷⁰ This is discussed in the next section.

To start with rule-based decisions, on a very high level, explanation is instrumental to ideals and principles of transparent governance. E.g. in functioning democracies, the aim of empowering subjects (citizens) with knowledge about public decision making rests on assumptions that this lets them understand how they are governed, challenge such decisions, and inform their electoral vote.⁷¹ The point is to counter the abuse of power. These ideals and principles are recognizably alive in the background of rules that see to the explanation of individual decisions. For example, in the demand that laws that decisions are based on are written in clear and understandable language; that decisions are accompanied with a statement about what particular legal provision(s) grounds the decision and on what grounds it can be contested.⁷² But the ideals also come much critiqued. E.g., in light of the complexity of contemporary states, of laws, and of governing institutions, naive notions of transparency are called out as simplistic and unable to produce the kind of understanding they are relied on for.⁷³ This in turn affects the regard for individual explanation demands discussed above. As we will see, the introduction of ever advancing technological methods in governance and decision making has amplified such critiques.⁷⁴

70 As we will see, the distinction is not always useful or tenable. In the medical domain for example, instrumental and intrinsic value tend to conflate: respecting patients in their humanity 'also' serves the instrumental aim of building the trust relationship that is instrumental to a successful treatment relationship. See Pierce on this particular point, Robin L. Pierce, 'Medical Privacy: Where Deontology and Consequentialism Meet', in *The Handbook of Privacy*, edited by Bart van der Sloot and Aviva de Groot (Amsterdam University Press, 2018).

71 David Heald, 'Variations of Transparency', in *Transparency: The Key to Better Governance?*, edited by Christopher Hood and David Heald (Oxford, UK: Oxford University Press, 2011).

72 Patrick Birkinshaw, 'Transparency as a Human Right', in *Transparency: The Key to Better Governance?*, edited by Christopher Hood and David Heald (Oxford, UK: Oxford University Press, 2011), 55. B.J. Koops, 'On decision transparency, or how to enhance data protection after the computational turn.', in *Privacy, due process and the computational turn*, edited by M Hildebrandt and K de Vries (Abingdon: Routledge, 2013), 169–220.

73 Mark Fenster, 'Transparency in Search of a Theory', *European Journal of Social Theory* 18, nr. 2 (1 May 2015): 150–67.

74 Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability', *New Media & Society* 20, nr. 3 (1 March 2018): 973–89; Joanna Bryson, 'The origins of bias and the limits of transparency'; Jakko Kemper and Daan Kolkman, 'Transparent to whom? No algorithmic accountability without a critical audience', *Information, Communication & Society* 22, nr. 4 (2019).

Sticking with typical public rule-based decisions for a bit longer, on the local level, an explanation of a decision about an individuals' welfare application typically clarifies the individual's situation's 'fit' with the policy at hand. The policy is based on general (welfare) laws, and itself consists of more precise eligibility rules. The translation of the person's situation in terms of these rules, and in light of additional rules and principles such as the avoidance of administering 'disproportionate hardship' needs to be justified. In a judicial motivation of a criminal sentence, the translation is of the behavior of the decision subject in terms of the legally described offense the person is charged with, and additional rules and principles such as those about culpability. These examples show how an individual explanation of a decision to some extent also explains the rules that govern that decision. The translation would not make sense otherwise. Explaining how an individuals' situation fits a rule, entails to generalize the person's situation to the level *of* the rule. In the words of Schauer, "the central point is that to say 'x because y' is not only to say x, but to say y as well."⁷⁵ And in light of how there are theoretically uncountable ways to argue such generalizations, the 'guiding morality' of explainers is a factor of influence on their reasoning, and so, it is of interest.⁷⁶

As we will see, the extent to which 'y' is 'said' varies considerably. This influences the extent to which the process and outcome of a decision will be understandable. But how much *sense* a decision makes to explainees also depends on the extent to which 'y,' the rule, is not just said, but explained, and even justified, in societal terms. Widdershoven writes how answers to *this* 'why' (put differently: 'why 'y'') belong to the realm of legal principles.⁷⁷ The thesis especially engages with this point, or more precisely with what, then, can be expected of such principles and how that reflects back on their own instructive clarity and 'optimal' measure of codification. E.g., terms like 'sexual abuse' in law have seen progressive explanations so that previously excluded behavior now qualified for the label (such as marital sexual abuse.) Before that, a person would be explained that what they experienced was not that. This introduces an important theme of the thesis, which is that explanations are also 'knowledge making moments,' in which conclusions about earlier knowledge making moments (laws, rules, and concepts used in them) are negotiated. This is not typically named as an instrumental explanation aim, but some sense of this view does express in arguments for explaining *per se*, as the next section will show.

The introduction of this theme also helps to understand the pertinence of explanation rules as decisions about what is of interest to know, and what needs to be justified. Another way to think about this is to understand the example of the legal qualification of abuse as an agreed upon diagnosis of a social situation and a type of behavior. This phrasing, in turn, helps to understand a similarity between rule-based decisions and expertise-based decisions such as *medical* diagnoses, and the explanation rules that see to such decisions.

⁷⁵ Frederick Schauer, 'Giving Reasons', *Stanford Law Review* 47, nr. 4 (April 1995): 633–59.

⁷⁶ Cohen, 'The Ethical Basis of Legal Criticism', 216.

⁷⁷ Rob Widdershoven, 'Een ervaring als staatsraad-generaal: op zoek naar een rechtsbeginsel', in *De conclusie voorbij. Liber amicorum aangeboden aan Jaap Polak*, edited by M Bosma e.a. (Ars Aequi, 2017), 89.

Sometimes, the two kinds of decisions are gathered under the same explanation rule. Diagnoses and similar expert opinions enter courtrooms (e.g., to establish culpability), but also administrative decision practices (e.g., to establish asylum seekers' risk of persecution), as evidence about a person's state. This way diagnoses come to ground further decisions such as those on a perpetrator's culpability or an applicant's request with an administrative body. The 'rules' to explain here could refer to the diagnostic concepts, social conceptualizations and so on, but this is not necessarily what explanation rules oblige to do.⁷⁸ Within the medical domain itself, the most well-known instrumental explanation aim was already named: to gain a patient's informed consent before diagnostics and treatment are engaged with. As we will see, the word 'explanation' is not used in the main law that establishes this right in The Netherlands. The law's main instrumental aim of 'informing' patients is to allow them to make informed treatment choices. Chapter 5 discusses how this focus is criticized for downplaying other values including, importantly, the intrinsic value of explanatory exchanges.

2.1.3 Dignitarian aims and arguments

This section introduces aims of explanation that are considered to be fundamentally important, sometimes argued as less directly instrumental. Fundamental importance of understanding the hows and why's of decisions is commonly argued on human rights-based understandings of autonomy and/or self-determination;⁷⁹ a less instrumental argument holds that respecting persons' basic needs for understanding their environment follows from their humanity.⁸⁰ When such fundamental, including less instrumental aims are seen to be in peril, mentions of 'dignity' are never far away;⁸¹ the arguments are also known as 'dignitarian.' The concept of dignity is much referred to in AI governance instruments and initiatives, not least those concerned with (rights to) explanation. But the concept of dignity and its value for the well-being of humanity in governance instruments also attracts critique. Various arguments used in these debates (about reasoning, about humaneness, and relational dignity/autonomy) play an important role in the thesis. This section therefore discusses it at some length.

78 Although these subjects are inevitably also what the decision sees to, which regularly leads to judicial conflicts. The Dutch Immigration Agency is repeatedly under fire for using wrongful methods to estimate asylum seekers' claims about their sexual orientation, and a recent court case led to an exceptional statement in the newspapers by psychiatrists who claimed a suspect was wrongly found culpable after judges dismissed their explanation of how 'rationally planned acts' were part of the suspects' psychosis and did not prove his sanity. Sabine Jansen, 'Trots of Schaamte? Het vervolg' (COC Nederland, March 2022); "Rechters zetten onze deskundigheid weg als een mening", *NRC*, 17 October, <https://www.nrc.nl/nieuws/2022/10/17/rechters-zetten-onze-deskundigheid-weg-als-een-mening>-a4145406.

79 Within the European regulatory space typically accompanied by references to Kant and the need to see humans as ends in themselves, which will be discussed critically a bit further on EDPS Ethics Advisory Group, 'Towards a digital ethics', 2018.

80 Bygrave, 'Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions'.

81 To quote Foster: "Burrow beneath any right worth defending and, I contend, you will hit dignity." Foster, *Human Dignity in Bioethics and Law*, 17.

Writings on human dignity cover a range of aspects: what it is; what it follows from; what and who it applies to; and whether the concept can usefully, and (universally) beneficially inform decision making and explanation. To start with the level of definition: there is no agreed upon definition of human dignity. The many laws that use the concept describe it as either a state, a right, an inherent or inviolable trait, an objective, a value, a foundational notion, or leave it undefined altogether.⁸² This disparity persists in descriptions of what dignity demands, or commands, in these laws: to be respected, protected, secured, reaffirmed, or recognized. Overall, dignity's instrumental usefulness

82 The United Nations Charter expresses the determination to save future generations from torments like those experienced in the First and Second World Wars and "to reaffirm faith in fundamental human rights, in the dignity and worth of the human person." The UN Universal Declaration of Human Rights holds that "all human beings are born free and equal in dignity and rights," and ICCPR and ICESCR state that "recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world." Rights "derive from the inherent dignity of the human person." ICESCR's article 13 (on education) determines that "education shall be directed to the full development of the human personality and the sense of its dignity." UNESCO's Universal Declaration on the Human Genome and Human Rights in article 11 forbids "[p]ractices which are contrary to human dignity, such as reproductive cloning of human beings." The African (Banjul) Charter on Human and People's rights declares that "freedom, equality, justice and dignity are essential objectives for the achievement of the legitimate aspirations of the African peoples." Background is added: "[c]onscious of their duty to achieve the total liberation of Africa, the peoples of which are still struggling for their dignity and genuine independence..." and explained: "[e]very individual shall have the right to the respect of the dignity inherent in a human being and to the recognition of his legal status. All forms of exploitation and degradation of man, particularly slavery, slave trade, torture, cruel, inhuman or degrading punishment and treatment shall be prohibited." The American Convention on Human Rights also relates dignity to (the prohibition of) slavery. About people's treatment in captivity or during forced labor (as a form of legal punishment) it states how that "shall not adversely affect the dignity or the physical or intellectual capacity of the prisoner." Article 11, on privacy, starts with "[e]veryone has the right to have his honor respected and his dignity recognized." The Oviedo Convention speaks of "protection of dignity and identity," and notes that the misuse of biology and medicine may lead to acts that endanger human dignity. Its Additional Protocol prohibits the conduction of research contrary to human dignity. It declares that cloning "genetically identical human beings" is an instrumentalization contrary to human dignity and thus constitutes such a misuse. The Declaration of Helsinki establishes ethical principles for research involving humans. It places protection of dignity in the hands of the researching physicians, and the concept in line with life, health, (dignity), integrity, right to self-determination, privacy and confidentiality. Notwithstanding the promotion of dignity as a European foundational value, the concept is missing in the main body of the Council of Europe's (CoE) European Convention on Human Rights (ECHR from hereon). Aside from the recognition of dignity as featured in the UDHR, the concept is only named in Protocol no. 13's prohibition of torture and other inhumane or degrading treatment: "...abolition of the death penalty is essential for the protection of this right and for the full recognition of the inherent dignity of all human beings." Dignity as a foundational value was primarily developed by the European Court of Human Rights (ECtHR), which established that "respect for human dignity and human freedom is "the very essence of the Convention." The CoE's Modernized Convention 108 (108+) on data protection reiterates how it is "necessary to secure the human dignity .. of every individual." In the European Union Charter of Fundamental Rights (the Charter hereafter), the first of seven titles is "Dignity" and its first article ("Human Dignity") states that "[h]uman dignity is inviolable. It must be respected and protected." In the Charter's preamble it holds the EU is founded on the "indivisible, universal values of human dignity, freedom, equality and solidarity." Mentions of dignity can be found in other EU documents, for example as the "bedrock" of the Charter or as playing a "foundational role for it. In the GDPR, dignity is only named in the context of labor relations: Member States may enact their own data processing rules for employees as long as they build in "suitable and specific measures to safeguard the data subject's human dignity, legitimate interests and fundamental rights."

in securing freedom-from-maltreatment rights (saliently, freedom from torture) is better developed than more positive applications of what humane treatment should entail,⁸³ including what it means for the right & duty to explain.

One reason for this underdevelopment arguably follows from the fact that we don't universally agree on what 'humane' treatment is, who deserves it, and on what grounds. From its birth onward, Law is infamous for its iterative oppressive delineations of who counts as a rights-bearing subjects whose dignity needs recognition (the enslaved, women, people of color, indigenous groups and many more).⁸⁴ Medical ethicist and health law scholar Foster is critical of the choice of scholars, lawmakers, and bioethicists to derive fundamental human needs and values from what makes people human: a biological state, trait, or nature; especially 'the ability for rational thought.'⁸⁵ Such choices enable to fundamentally *disrespect* those who can't, or aren't considered to exhibit the state, trait, or nature. It grounds the ascription of 'inhuman' natures to groups of persons on the basis of physical, cultural or other characteristics. This kind of 'dignity reasoning' has historically let those with decisional power exclude whole groups from the so-called shared value space, and therewith from basic rights, 'dignified' treatment, and participation in decision making.⁸⁶ Therewith it stands in the way of the development of normative notions about human nature (positive applications of dignity) that don't also produce notions about inhuman nature.⁸⁷ Critical authors therefore argue to ground notions about what humans fundamentally need, and what values to uphold for them, on their *humaneness*: a positive understanding of humanness. Foster describes it as "that in whose absence people can live but not thrive," that what relies on, and requires, mutual care.⁸⁸ Dignity in this view is upheld through human moral relationships, and in the relationships of humans with themselves. I.e., when one disrespects an other's dignity one also disrespects oneself as a participant in dignified humanity.⁸⁹

83 The concept is used to label certain rights as non-negotiable, and to mark the extent to which restrictions to human rights are legally permitted: dignity is at its best as 'restrictions restrictor' Max Vetzo, Janneke Gerards, en Remco Nehmelman, *Algoritmes en Grondrechten*, Montaigne reeks (Boom Juridisch, 2018); That said, even such establishments necessarily need to be understood to be a promise, rather than a practice Mary Neal, 'Respect for Human Dignity as "Substantive Basic Norm"', *International Journal of Law in Context* 10, nr. 1 (March 2014): 43.

84 Ernst Hirsch Ballin, *Advanced Introduction to Legal Research Methods* (Edward Elgar Publishing 2020) 27.

85 Foster, *Human Dignity in Bioethics and Law*, ch1 and 2.

86 McCrudden, 'Human Dignity and Judicial Interpretation of Human Rights'.

87 Linda Martin Alcoff, 'Philosophy and Philosophical Practice: Eurocentrism as an Epistemology of Ignorance', in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017); Foster, *Human Dignity in Bioethics and Law*.

88 Foster cites Nussbaum, Dworkin, Hale, Debes and others who agree that dignity is necessarily and inherently shared, Foster, *Human Dignity in Bioethics and Law*, 95.

89 Neal, 'Respect for Human Dignity as "Substantive Basic Norm"', 29.

In writings about what dignitarian demands should mean for individual treatment, (more) individual versus (more) relational views on humaneness co-exist. The first expresses in respect for individual autonomy as ‘freedom from’ other humans’ maltreatment, and/ or freedom ‘to’ make their own choices (negative and positive freedom). Critics argue that this sketches a too individualistic picture of human functioning and ignores how people are inter-dependent nodes in social networks.⁹⁰ This understanding of autonomy, and dignity, as relational requires that people are protected in terms of how they are dependent on each other, and on larger societal safeguards and affordances.⁹¹ More positively put, to ensure ‘freedom for’ people to co-exist in safety and equality.

These discussions of humane-ness v. human-ness, and individual v. relational autonomy/dignity, can and do inform explanation-relevant notions, rights and duties. They for example allow to take the ability of humans to reason or ‘think rationally’ into account in a more responsible way: as one expression of human thriving that includes many other forms of ‘cognitive activity’ (remembering, knowing, daydreaming) that follow from (self)awareness.⁹² As Medina writes, “meaning-making and meaning-sharing are crucial aspects of a dignified human life.”⁹³ But it remains important to prevent the use of such arguments as causal (‘we thrive because we reason’), thereby inviting exclusionary definitions of what qualifies as reason.⁹⁴ We should also not too easily assume respect for reason is served. The argument that “[I]ndividuals whose lives are governed by law are treated by it as thinkers, persons who can “grasp and grapple with the rationale of that governance,” which reveals “an implicit commitment to dignity in the tissues and sinews of law”⁹⁵ invites critical questions about how laws do not necessarily do this for everyone.

90 ‘the most accurate description of an individual will be in terms of the nexus of relationships in which she exists,’ Foster, *Human Dignity in Bioethics and Law*, 12.

91 See for example Rouvroy, arguing that human autonomy is contingent on socio-economic, educational and other factors Antoinette Rouvroy, “‘Of Data and Men’. Fundamental Rights and Freedoms in a World of Big Data.’ (2016) T-PD-BUR(2015)09REV Council of Europe, Directorate General of Human Rights and Rule of Law. 54 and further.

92 As for example expressed by Lorraine Code, “[a]lthough I do not think there is an essential ‘humanness,’ I do think cognitive activity is so central to human life that any evaluation of human character must take the quality of this activity into account (...) perceiving, remembering, reasoning, knowing, believing, speaking, imagining, daydreaming; activities that have their source in experience of the world and of oneself as part of the world: in awareness and self-awareness.” Code, *Epistemic Responsibility*, 52.

93 José Medina, ‘Varieties of Hermeneutical Injustice’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017), 41.

94 Which has for example been implicit in European philosophy, resulting in reason itself being ‘in need of justification’ Lewis Gordon, *Freedom, Justice, and Decolonization* (Routledge, 2020).

95 Jeremy Waldron, ‘How Law Protects Dignity’, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 15 December 2011), 211, 222, <https://papers.ssrn.com/abstract=1973341> as cited by Neal (2014) p.40.

In contemporary perceptions of the AI-induced explanation crisis, especially of the individual right to explanation, relations between dignity and the human ability to reason are much mentioned. The ability to reason for example features in the European understanding of autonomy ‘in the Kantian tradition,’ as was mentioned:⁹⁶ people need to be respected as able and responsible, individual moral reasoners. Such individuals deserve to be faced by decision makers who bear arguments, and they deserve the freedom to make up their own minds. But critics explain how Kant himself excluded whole groups of people as able reasoners (notably: women),⁹⁷ and European colonialist ‘traditions’ dismissed whole communities’ ability to reason, the right to explanation of decisions about them, and the freedom to make up their own minds.⁹⁸

One explanation for the European region’s strong notion of individual autonomy as in need of ‘dignitarian respect’ is seen to lie in the dehumanizing deeds of the Nazis in (and around) the Second World War. Put simply, the objectification and subsequent extermination of Jews, Roma, Sinti, those perceived to be mentally or physically disabled, homosexuals and other groups on the basis of (attributed) group characteristics required the erasure of their individuality. Innovative (analog) automation on the basis of personal information is seen to have played an important role in facilitating this treatment, which is used as a warning “not to treat people as mere aggregates of data.”⁹⁹ Leta Jones argues how this explains the foundational importance that European data protection laws have given to “a human in the loop” of decisional processes,¹⁰⁰ a human being who can justify (explain) their actions *to* ‘data subjects’ and be held accountable *by* them. The role of data protection law in establishing a fundamental right to individual explanation of automated decisions is seminal (as will be discussed some more below). But the framework’s reliance on individual rights, duties, capabilities, and autonomous understanding are also critiqued as weak points, as was discussed in the introduction. This thesis contributes to these discussions by clarifying what it thinks is fundamentally important about the practice

96 EDPS Ethics Advisory Group, ‘Towards a digital ethics’.

97 Pauline Kleingeld, ‘On Dealing with Kant’s Sexism and Racism’ 2, nr. 2 (2019): 21.

98 Alcoff, ‘Philosophy and Philosophical Practice: Eurocentrism as an Epistemology of Ignorance’; Rebecca Tsosie, ‘Indigenous Peoples, Anthropology, and the Legacy of Epistemic Injustice’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017).

99 EDPS Ethics Advisory Group, ‘Towards a Digital Ethics’ (2018) 17. “Human dignity as the foundation of human rights implies that meaningful human intervention and participation must be possible in matters that concern human beings and their environment. Therefore, in contrast to the automation of production, it is not appropriate to manage and decide about humans in the way we manage and decide about objects or data, even if this is technically conceivable. Such an ‘autonomous’ management of human beings would be unwelcome, and it would undermine the deeply entrenched European core values.” The European Group on Ethics in Science and New Technologies (EGE) | EGE - Research and Innovation - European Commission, ‘Statement on Artificial Intelligence, Robotics and “Autonomous” Systems’, 9–10; see also, EDPS Ethics Advisory Group, ‘Towards a digital ethics’, 17.

100 Jones, ‘The Right to a Human in the Loop’.

of explaining, and how that should inform our governance of it. It makes its own, relational, ‘dignitarian argument.’

2.2 Perceptions on a right in crisis

2.2.1 Decisions are not made here anymore: a crisis of the individual right

The phrase “Computer says no,”¹⁰¹ a citation from a 2004 TV sketch, has come to represent situations in which explainers hide behind their computers: situations of unreasoned, and therefore unreasonable decisional authority. But our contemporary ‘computers’ are not the same. Especially machine-learning based algorithmic tools that are used in decision making are increasingly seen to challenge how explanation rights of people can be served at all. Different reasons are contained in such arguments, related to technological characteristics of computation, socio-technical dimensions and questions of human machine interaction, commercial and organizational factors.¹⁰² Various notions will feature in the following sections. The first two discuss perceived challenges to the individual rights of explainees as understanders of, and as participants in, decisional processes. The one after proceeds to the duties of explainers, and the last explains how these duties are drawn into doubt in perceptions on the concept and value of explanations as we know them.

To start with, the explosion of data that are made available for use in modern decision support technologies has fueled methods of knowledge and decision making that go beyond what could be done in the past. Predecessors of modern computational systems were already built to support, and to accelerate human practices of ‘sorting’ people and phenomena to inform decision making.¹⁰³ Such sorting was done based on data that humans had gathered and had labeled as informative. The further development of these technologies was driven by mid-20th century imaginations of how concepts of sorting itself could be ‘improved’: how machines could enable new modes of analysis. It fueled a quest for observable ‘traces’ of events and human behaviors. These would yield patterns: correlative relations that were imagined, in turn, to yield predictive insights about human behavior. Such insights would allow the making of targeted

101 Little Britain, television series BBC 2004

102 E.g., by asking machines to reason in ways that humans cannot, ‘inexplicability’ as we know it is the point James Larus et al, ‘When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making’, White paper (Informatics Europe & EUACM, 1 January 2018), <https://doi.org/10.1145/3185595>; Neyland takes on lazy accounts of obscurity by engaging with the socio-technical Daniel Neyland, ‘Bearing Account-Able Witness to the Ethical Algorithmic System’, *Science, Technology, & Human Values* 41, nr. 1 (January 2016): 50–76; Gürses and Hoboken discuss challenges that follow from the networked, real-time, software-as-a-service infrastructures Seda Gürses and Joris van Hoboken, ‘Privacy after the Agile Turn (Version 3)’, *Open Science Framework*, 2017 <https://osf.io/preprints/socarxiv/9gy73/>.

103 Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences*, Revised edition (Cambridge, Massachusetts London, England: The MIT Press, 2000).

policy, and allow to profile citizens for multi-fold public and private purposes.¹⁰⁴ Today, automated decision making (support) systems (ADS) produce behavioral and diagnostic predictions, assessments, recommendations and decisions that are used to anticipate, manipulate, support and govern human states and behavior.

Not nearly all applications are as advanced as they were imagined to be, and their potential is frequently overrated as well as misunderstood.¹⁰⁵ But challenges to individual explanation rights don't 'correlate neatly' to technological complexity. They already arise in relatively simple settings. E.g., administrative bodies in The Netherlands and elsewhere have been connecting previously separate databases of basic registrations (think addresses, vehicles, income, household composition). Such databases notoriously contain 'dirty data': incomplete, incorrect, inaccurate or obsolete items of information. The use of such data leads to incorrect decisions and explainees have a hard time understanding how that happens. And even when collected data are able to support correct in-context conclusions (decisions), when such conclusions travel out of context, i.e. are incorporated by other administrative bodies as components of *their* decisions, out-of-context interpretations led to surprising mistakes. The algorithmically driven analytic systems used for such decision making may be technologically explainable but finding earlier decisional steps that led to the unwanted outcome still becomes near-impossible—for explainees, decision makers, and even for judges that explainees turn to as a last resort.¹⁰⁶

Complementary to 'analog' legal protections against wrongful decisions in a decisional domain's laws (e.g., Criminal and Administrative Law), European data protection laws have played an important role in strengthening the position of explainees as data subjects with data subject rights. By now, the right to data protection is also recognized in the EU's Charter of Fundamental Human Rights, and the Council of Europe has its Modernized Convention 108. Data protection has been described as an 'enabling' right,¹⁰⁷ enabling explainees several interventions that are meant to (at least) reduce (effects of) information inequality,¹⁰⁸ and at best protect from inhuman, possibly inhumane treatment by machines. These laws themselves have become complex to

104 Jill Lepore, *If Then: How the Simulmatics Corporation Invented the Future*, Illustrated edition (New York: Liveright, 2020).

105 Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* (The MIT Press, 2018).

106 Marlies van Eck, 'Geautomatiseerde ketenbesluiten & rechtsbescherming: Een onderzoek naar de praktijk van geautomatiseerde ketenbesluiten over een financieel belang in relatie tot rechtsbescherming.' (Doctoral Thesis, Tilburg University, 2018).

107 Manon Oostveen and Kristina Irion, 'The Golden Age of Personal Data: How to Regulate an Enabling Fundamental Right?', in *Personal Data in Competition, Consumer Protection and IP Law - Towards a Holistic Approach?*, edited by Bakhoun et al (Berlin: Springer, 2017).

108 Lokke Moerel and Marijn Storm, 'Automated Decisions Based on Profiling - Information, Explanation or Justification? That Is the Question!', *Oxford Law Faculty, Law and Autonomous Systems Series* (blog), 27 April 2018, <https://www.law.ox.ac.uk/business-law-blog/blog/2018/04/law-and-autonomous-systems-series-automated-decisions-based-profiling>.

navigate, however. The contemporary European data protection regime is described as a “potpourri of disparate rights and principles that may more usefully be explained as a “fundamental right to having a (set of) rule(s) regulating the processing of personal data,” than as set of clear-cut solutions.¹⁰⁹

With regard to the right to explanation, the EU’s General Data Protection Regulation (GDPR) has led the way since it entered into force in 2018, replacing the Data Protection Directive. The GDPR’s explicit positioning of the principle of transparency, in combination with other updates, meant to create a deeper, broader and stronger accountability regime for ADM.¹¹⁰ This is seen as an expression of the importance the regime attaches to data subjects’ knowledge and understanding of “risks, rules, safeguards and rights”¹¹¹ with regard to personal data processing. The Regulation updated subjects rights of access, correction, and addition of information; of contestation, and, as stated in Recital 71 and later confirmed the European Data Protection Board, established a right to explanation.¹¹² Most explicitly in the law subjects have a right to “meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing.”¹¹³ The provisions have led to an explosion of research about what, if anything, constitutes meaningful information.¹¹⁴ The jury’s still out. At the time of writing, the Vienna Regional Administrative Court put the question to the European Court of Justice.¹¹⁵

The European Data Protection Board’s Guidelines on Transparency explain that the aim of GDPR’s provisions is to “*meaningfully position* [data subjects] so that they can vindicate their rights, and hold data controllers accountable for the processing of

109 Lorenzo Dalla Corte, ‘A Right to a Rule: On the Substance and Essence of the Fundamental Right to Personal Data Protection’, in *Data Protection and Privacy: Data Protection and Democracy*, edited by D. Halliman et al. (Hart Publishing, 2020).

110 Kaminski, ‘The Right to Explanation, Explained’.

111 European Commission, ‘Stronger protection, new opportunities: Commission guidance on the direct application of the General Data Protection Regulation as of 25 May 2018, COM(2018) 43 final’, 24 January 2018, 11. See also GDPR recital 39.

112 Article 29 Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP 251rev.1)’, 27.

113 This is also true when no personal data are collected to make the decision, as long as it (legally or similarly significantly) affects the subject directly. The fact that articles 13&14 pertain to cases of data collection is therefore no ‘escape’ for controllers.

114 This topic is discussed further on. For now, see e.g., Moerel and Storm, ‘Automated Decisions Based on Profiling - Information, Explanation or Justification? That Is the Question!’; Aviva de Groot and Sascha van Schendel, ‘Explaining Responsibly: a panel discussion with Reuben Binns, Michael Veale, Martijn van Otterlo, and Rune Nyrop’ (Tilting Perspectives, Tilburg, 2019), <https://easychair.org/smart-program/TILTING2019/2019-05-17.html#talk:89359>; and Kaminski’s explanation of these discussions for challenged US scholars, Kaminski, ‘The Right to Explanation, Explained’.

115 Landesverwaltungsgericht Wien, VGW-101/042/791/2020-44 (Landesverwaltungsgericht Wien 11 February 2022); As referred to in ‘Automated Decision-Making Under the GDPR - A Comprehensive Case-Law Analysis’ (Future of Privacy Forum, May 2022).

their personal data.”¹¹⁶ But whether individuals can realistically occupy such a position is much questioned, and what amounts to accountable use of computational systems has become a field of research in itself. Authors describe amalgams of humans and globally networked computational systems that together make up a vast and highly complex ‘socio-technical space.’¹¹⁷ On such a view, ‘meaningful explanations’ may need to “encompass, amongst others, the algorithmic system’s reason for existence, the context of the development, the effects of the system.”¹¹⁸ In other words, explanations of what happens at levels far above and beyond the individual, local level. That conclusion seems inescapable in light of how much knowledge that informs ADM systems is ‘aggregate knowledge,’ and effects play out on group levels to the extent that individual treatment becomes hard to identify, and therewith challenge, because these are prerequisites for most explanation and (other) legal protection rights.

2.2.2 Eroding possibilities for meaningful subject participation

The first European Data Protection proposals already expressed concerns that decision subjects would cease to be able to participate in decision making when decisions become based on their “data traces.”¹¹⁹ Such concerns have only deepened. Public Administrations are typically on the forefront with regard to the implementation of novel data science methods, and those over whom they exert the most power are the first to be affected.¹²⁰ Eubanks and others convincingly researched how the impossibility for citizens in need of support to understand and responsibly interact with the many ADM systems used in decision making about them erodes their “feelings of competence and proficiency,” exhausting them into a loss of autonomy, (self-)respect and dignity.¹²¹ But in commercial contexts as well, subjects are said to unwittingly participate in decision making. Everything they do generates data that is captured in increasingly less trace-able ways, and combined with “traces/observations from other sources” unrelated to them.¹²² The correlative predictions this produces are acted on in anything from creditworthiness scores to job applications to insurance

116 Article 29 Working Party, ‘Guidelines on Transparency under Regulation 2016/679 (wp260rev.01)’, 26 (original emphasis).

117 Adam Greenfield, *Radical Technologies: The Design of Everyday Life* (London ; New York: Verso, 2017); Gürses and van Hoboken, ‘Privacy after the Agile Turn (Version 3)’; Edwards en Veale, ‘Slave to the Algorithm?’

118 Wieringa, ‘What to account for when accounting for algorithms’, 7.

119 ‘Explanatory text for Proposal for a Council Directive concerning the protection of individuals in relation to the processing of personal data’ (Commission of the European Communities, 1990), 29.

120 Lepore, *If Then*; Black, *IBM and the Holocaust: The Strategic Alliance Between Nazi Germany and America’s Most Powerful Corporation*; Benjamin, *Race After Technology*; Broussard, *Artificial Unintelligence*.

121 Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, 1st edition (St. Martin’s Press, Macmillan Publishers, 2018), 194; Khiara M. Bridges, *The Poverty of Privacy Rights* (Stanford University Press, 2017).

122 Seda Gürses and Joris van Hoboken, ‘Privacy after the Agile Turn (Version 3)’ 2017 <https://osf.io/preprints/socarxiv/9gy73/>.

fraud, depending on the legal context. As Malik writes, “this effectively rewards and punishes people” for things over which they have (at least in part and depending on the data point, entirely) no control.¹²³ And while he rightly points out that doing people injustice on the basis of circumstances that they have no control over is not a new phenomenon, the ways in which technology is engaged to identify and select such circumstances, is.

Current computational methods such as Machine Learning no longer serve decision makers like statistical, hypothesis driven techniques did, even if those methods have been absorbed.¹²⁴ The former were designed to support and further (theoretically) explainable notions of what was considered to be good, bad, interesting, or optimal. But this furthering is being ‘outsourced.’ To illustrate, Amoores uses the example of a ‘riot:’ a type of gathering that is considered as a threat by a government, and that they therefore want to recognize and respond to. Ideally, before they happen, which is what they expect ADM systems can help them with. Such governments now have machine learning methods at their disposal that identify human movements and gatherings as possibly riotous. Based on excessive data crunching, systems classify, identify, predict, and recommend courses of action: they make and support decisions. Rather than (just) using existing human labels and hypotheses, systems are asked to create their own ground truths, their own decision rules.¹²⁵ In the process, they get to influence and change definitions of ‘riot’ versus non-violent protest, and versus other group expressions.¹²⁶ The more unstructured the signals from the world are that such systems are allowed to train on, the more choices about what to take into account, and what points or patterns to discard are inevitably made in the process. And where human rule makers (lawmakers) can be made to justify their choices, this information can simply not be gotten from machine learning algorithms.¹²⁷ Next to the more obvious ‘right to explanation’ concerns this raises, authors have started to warn how the individual and communal rights of persons to a basic understanding of their environs, of how the world around them works, is in peril.¹²⁸ Some argue that human dignity requires that persons’ understanding, participation, and “co-decision making” rights should be met even if an AI-generated outcome is “legally and morally acceptable.”¹²⁹ In other words, these are arguments for ‘explanation *per se*.’

123 Momin M. Malik, ‘A Hierarchy of Limitations in Machine Learning’, *ArXiv Preprint ArXiv:2002.05193*, 29 February 2020, 29, <http://arxiv.org/abs/2002.05193>.

124 Malik, ‘A Hierarchy of Limitations in Machine Learning’.

125 Louise Amoores, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others* (Duke University Press, 2020).

126 Such algorithms ‘establish new patterns of good and bad, thresholds of normality and abnormality against which actions are calibrated’ Amoores, 6.

127 Amoores, 111–12.

128 Bygrave, ‘Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions’.

129 Which reads a bit like a contradiction in terms, Nathan Colaner, ‘Is Explainable Artificial Intelligence Intrinsically Valuable?’, *AI & Society* 37, nr. 1 (1 March 2022): 231–38.

To conclude these two sections; messy combinations of high-tech, low-tech, and no-tech (human discretion) methods are keeping researchers and explainees alike occupied with finding out what happens in decision making. Explanation rules in place have clearly not prevented the erosion of applicable rights and principles – but what about the explainers who faced these explainees? *Could* they have done a better job? The next section discusses how their abilities are seen to be challenged by technology, too.

2.2.3 Explainers are challenged to fulfill their duties

Vredenburg argues that ‘workers’ need to understand the social systems they are employed in if we want to rely on them to endorse, reject, or oppose possibly perverted or wrongful courses of their institutions.¹³⁰ This argument for the intrinsic value of Explainable Artificial Intelligence (XAI) advances the meaningful participation concerns of the previous section to the plights of explainers. The author names the teacher whose school system turns out to be designed so as to uphold unwelcome class structures, rather than promote equal education opportunity. As an important type of explainer, the roles of teachers will be engaged with again later on in the thesis.

The teacher’s burden is deliberate. Arguably, institutionalized decision makers and explainers are expected to engage critically with whatever methods support their decisional processes, and with the knowledges that ground their decisions in terms of law, policy, science, history, or all of these, depending on their context. If they were not, we could let machines do their jobs. At the same time they are relied on to uphold systems that are politically agreed on: to abide by agreed upon aims, rules, and methods. Needless to say, this creates tensions. A recent reflection by a Childcare Benefits Scandal judge illustrates such an instance. “There is also a more abstract question to raise,” he wrote. “What do you do when the lawmaker wants ‘A’, but that conflicts with everything that you feel, think, and know. Are you allowed to, are you brave enough to decide ‘B’? And what’s the use, if the appeals courts tells you later it is ‘A’ after all?”¹³¹ We may wonder why the judge calls the question ‘abstract.’ Very real and recent history teaches what can happen if judges and other system workers cease to question their conscience.¹³² Our ideas about what is right are under perpetual development, and so we need to be able to question and debate about what informs our decisions.

So long as we have humans employed in important explanation roles, this argument is convincing. Still there is arguably ‘competition’ between arguments for individual explainers as the final providers of humane resolutions, and for dealing with (some) AI-infused explanation problems, and solutions, in ways that allow to make more general

¹³⁰ Kate Vredenburg, ‘Freedom at Work: Understanding, Alienation, and the AI-Driven Workplace’, *Canadian Journal of Philosophy* 52, nr. 1 (9 February 2022): 3.

¹³¹ Judge van Rijn, reflecting, ‘Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken’.

¹³² Arendt, *Responsibility and Judgment*.

justificatory statements.¹³³ One obvious argument for the latter is that explainers are generally not trained in AI. They won't be able to review a calculated outcome if they aren't able to 'do the math' that their tools do for them. But perhaps, no-one can. As Larus put it bluntly, "for many ML models, in particular deep neural nets, inexplicability is fundamental."¹³⁴

Taking a step back, for critical questioning to remain possible, our choices and decisions should at least be question-able, which at least means: not hidden.¹³⁵ The previous sections discussed how choices made by and in contemporary decision support systems are said to be precisely that: hidden, inaccessible, at least in ways that they were not so before. The problems this creates for explainers can perhaps be usefully understood as 'distancing problems.' To responsibly engage their moral judgment and ask critical questions, explainers need to maintain a responsible distance *vis-à-vis* their domain's rules and methods (the earlier 'tensions.'). And they are currently said to be either too much part of, or too far removed from where crucial choices in decision making happen to be able to do that.

To start with some 'too much distance' arguments, Malik discusses how it takes "years of specialized training" for ML researchers to build up an intuitive feel for the extreme abstractions that define their models' functioning. Such intuitions are generally not shared by the people that these models are creating knowledge about, and generally also not by explainers who work with ML systems and who are expected to explain machine conclusions to decision subjects. The ML researchers in Malik's example struggles to bridge their type of quantified knowledge to the qualitatively oriented understanding of system users:¹³⁶ our explainers, in explainee roles. In light of such challenges, some writers have argued that the dignitarian demand for individual explanation might require "modification," for example by explaining the oversight of a system rather than a system's internal functioning.¹³⁷ Others explore a 'flipped' burden of proof where commercial firms need to prove *ex ante* how their systems will produce

133 Andrew D. Selbst and Solon Barocas, 'The Intuitive Appeal of Explainable Machines', 87 *Fordham Law Review* 1085, 2018; Anderson identifies such competition between writers in the Epistemic Injustice domains, and argues that structural remedies enable individual virtuous behavior Elizabeth Anderson, 'Epistemic Justice as a Virtue of Social Institutions', *Social Epistemology* 26, nr. 2 (1 April 2012): 163–73.

134 Larus and others James Larus and others, 'When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making' (ACM 2018) 9 <https://dl.acm.org/citation.cfm?id=3185595> accessed 25 October 2018.

135 Roger Brownsword, 'In the year 2061: from law to technological management', *Law, Innovation and Technology* 7, nr. 1 (2 January 2015): 1–51.b

136 Malik, 'A Hierarchy of Limitations in Machine Learning', 8.

137 Deven R. Desai and Joshua A. Kroll, 'Trust But Verify: A Guide to Algorithms and the Law', *Harvard Journal of Law & Technology* 31, nr. 1 (27 April 2017): 44.

non-oppressive outcomes.¹³⁸ Yet others warn for the false sense of security that this may foster in explainees. They argue that individual explanations should aim to reveal the power structures at play in ADM systems, and for example demonstrate how, and not just that, a system was made immune to discriminatory bias.¹³⁹

The distance between those who are publicly tasked to make well-informed decisions, and those who effectively make those decisions, can become so large that Citron and Calo argue a ‘crisis of legitimacy.’ They see this in (US) Public Agencies that have outsourced the realization of their automation dreams for many decades.¹⁴⁰ In the process, they effectively outsourced the creation of expert domain knowledge to the point that civil servants are no longer able to explain (let alone justify) their decisions in court.¹⁴¹ In these and other domains, the number of explainers, “identifiable reasoning human subject[s],” who reveal a decreasing knowledge-ability of their methods is growing.¹⁴²

The honest denials of knowledge-ability that Citron and Calo’s civil servants displayed in court cited helps to reveal the underlying problem. But other writers argue that the possibility for explainer honesty itself is being undermined. In other words, they are no longer able to distance themselves sufficiently. In Europe, the GDPR demands that decision makers are able to perform meaningful oversight: that they have the authority, competence, and capability to analyze and change decisions, to consider “all the relevant data.” When a review is required of a decision that was fully automated, they are obliged to include any additional information that the data subject provides.¹⁴³ The provisions are not just inspired by earlier cited fears for automation’s objectifying distance, but by fears that decision makers will be *unable* to keep the necessary critical distance in the other direction: fears about the “objective and incontrovertible character [of machine conclusions] to which a human decision-maker may attach too much weight, thus abdicating his own responsibilities.”¹⁴⁴ This is an established problem. People are inclined to put (unfounded) faith in calculation, and to “defer to

138 Gianclaudio Malgieri and Frank A. Pasquale, ‘From Transparency to Justification: Toward Ex Ante Accountability for AI’, SSRN Scholarly Paper (Rochester, NY, 3 May 2022), <https://doi.org/10.2139/ssrn.4099657>.

139 Edwards and Veale, ‘Slave to the Algorithm?’, 65–66. See on the ‘consent fallacy’ Eoin Carolan, ‘The Continuing Problems with Online Consent under the EU’s Emerging Data Protection Principles’, *Computer Law & Security Review* 32, nr. 3 (June 2016): 462–73; Daniel J. Solove, ‘Privacy Self-Management and the Consent Dilemma’, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 4 November 2012).

140 Lepore, *If Then*; Calo and Citron, ‘The Automated Administrative State: A Crisis of Legitimacy’.

141 Calo and Citron, ‘The Automated Administrative State: A Crisis of Legitimacy’.

142 Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, 65.

143 Article 29 Working Party, ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP 251rev.1)’, 27.

144 ‘Amended proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data’ (Commission of the European Communities, 15 October 1992), 26.

the machine.”¹⁴⁵ Empirical, socio-technical research has convincingly revealed how human-machine interactions need to be studied up close if we want to understand how ADM tools influence human decision maker behavior.¹⁴⁶ Eubanks discussed an idealistic claim of a designer of an algorithmic system that produces child abuse risk scores. Their model was built to encourage critique: human intake screeners were to critically question the model’s output, to “undermine” it in that sense.¹⁴⁷ But that did not happen.¹⁴⁸ And even in instances where that does happen, ‘moral’ corrections are not necessarily made. US judges were found to ‘correct’ a recidivism risk prediction system more frequently for the qualifier ‘youth,’ and less for ‘colored,’ for example. With this, they consciously or unconsciously undermined something else: the purpose of the system to correct for their discriminatory biases in the first place.¹⁴⁹ In the (theoretical) absence of their own insight into how this happens, they won’t be able to justify their decisions—explanation comes in the form of (critical) scientific study.¹⁵⁰

The earlier cited mismatch between technological explainers and qualitatively trained (or inclined) understanders is not something that is easily resolved, if at all. This is particularly problematic for the responsible use of what are called ‘predictive systems.’ For one, Malik writes, prediction is a term inherited from statistics, but it had a different meaning there. This ‘orphan’ problem is exacerbated by the fact that especially, but not only lay people tend to impute ML models’ predictions with causality, even when no cause is ‘predicted’ to exist. Machine learning predictions are based on past performance of the system itself in terms of ‘correct’ correlations, which in no way comments on causal understandings or hypotheses. Still interpretations of ‘model predictive correctness’ are easily understood as assessments of predestination, and as ‘proven’ predictability of actual human behavior.¹⁵¹ A personal experience of this happened in a co-organized workshop in which the logic of AI’s predictive modeling was explained to members of a human rights lawyers’ network. The room exploded, with many questions from the audience testifying to fears of the absence of free will.¹⁵² Confusing dialogue about what ‘prediction’ means also took place between

145 Danielle Keats Citron, ‘Technological Due Process’, *Washington University Law Review* 85, nr. 6 : 1272.

146 Philip E Agre, ‘Toward a Critical Technical Practice: Lessons Learned in Trying to Reform AI’ in Geof Bowker and others (eds), *Bridging the Great Divide: Social Science, Technical Systems, and Cooperative Work* (1997); Andrew D Selbst and Solon Barocas, ‘The Intuitive Appeal of Explainable Machines’ [2018] 87 *Fordham Law Review* 1085.

147 Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, 168.

148 Also for political reasons: it was unclear to decision makers whether they would be seen to fail if they decided against the algo’s conclusion

149 Malik, ‘A Hierarchy of Limitations in Machine Learning’, 32.

150 Critical readers will respond that this is not a new problem – indeed it is not, as the next sections will discuss.

151 Malik, ‘A Hierarchy of Limitations in Machine Learning’, 22.

152 Personal experience while leading a workshop on ‘predictive’ AI with Sennay Ghebream, for the NJCM’s 45 years anniversary conference: “Onder populistisch vuur. Als zelfs de rechtsstaat niet meer vanzelfsprekend is,” 12 March 2020, De Balie, Amsterdam.

Judges and Government lawyers in what became a historical win over Government overreach in ADM implementation: the SYRi case.¹⁵³ That “the system does not actually predict human behavior” was argued by the defense even though the system was used to preemptively decide about humans in light of the systems predictions.¹⁵⁴

Such use subsequently creates instances that feed into a picture that comes to sustain the same false assumptions of causality: think of the over-policing of neighborhoods where crime is ‘predicted’ to happen, fraud detection directed at certain groups over others but also additional educational opportunities that get to be offered to those who are ‘predicted’ to do well.¹⁵⁵ Such practices are wrongful in more ways than one. The ‘created’ instances add to suggestive numbers that end up in reports and are publicly shared, feeding into a society’s discriminatory thinking. They add substantive contra-evidential burdens to people who are dependent on institutions that wrongfully distrust them, feeding into rightful distrust on their own part, with negative impact on their participation.¹⁵⁶ Because of this ‘confused use,’ and because it is hard to explain it away again, Malik suggests that using other terms instead of prediction, such as “back testing” might help to avoid confusion.¹⁵⁷

2.2.4 Challenges to typical explanation values, or: denial of crisis

The fact that conclusions of advanced ADM systems don’t rest on causal reasoning limits the *types* of ‘why’s that can be asked of them. An obvious complication (or at least, as we saw, confusing element) for various why’s that need to be justified in regulated explanation paradigms. Some argue that the enthusiastic uptake that such systems nonetheless enjoy, can be explained on the basis of changing notions about the merits of ‘causality’ as a concept of interest to inform decisions in the first place. Shifts of interest from *understanding* persons and situations, to *predicting* and/or

153 Personal case note, argument made by State attorney Pels Rijcken, Den Haag, 29 October 2019.

154 For an insightful discussion of AI’s correlative predictions, its problematic technological histories and contemporary consequences, see McQuillans’ first two chapters. ‘In the end,’ he writes, ‘the overarching correlation will be between the impacts of AI and the maintenance of existing social power, accompanied by the intensification of discriminative ordering.’ McQuillan, *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*.

155 Danielle Keats Citron and Frank A. Pasquale, ‘The Scored Society: Due Process for Automated Predictions’, *University of Maryland Francis King Carey School of Law Legal Studies Research Paper* 2014, nr. 8 (2014); Joanna Redden, Lina Dencik, and Harry Warne, ‘Datafied child welfare services: unpacking politics, economics and power’, *Policy Studies* 41, nr. 5 (2 September 2020): 507–26; Rashida Richardson, ‘Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities’, *Berkeley Technology Law Journal* 36, nr. 3 (2021); ‘We Sense Trouble: Automated Discrimination and Mass Surveillance in Predictive Policing in the Netherlands’ (Amnesty International, 2020).

156 Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*; Albertjan Tollenaar, ‘Bestuursrechtelijke normering en “big data”’, *Nederlands Tijdschrift voor Bestuursrecht* 2017, nr. 16 (2014): 133.

157 Malik, ‘A Hierarchy of Limitations in Machine Learning’, 26 & 45.

influencing persons and situations popularized different methods with which people and situations are studied.¹⁵⁸

Whether or not the shift of interest is intentional (or new, as the next section will discuss), Malik discusses four decisions that are at least implicitly made when ML methods are chosen, all of them consequential for what can be explained: the choice to use quantitative and not qualitative analysis; probabilistic modeling over other mathematical modeling or simulation; predictive modeling rather than explanatory modeling; and to rely on cross-validation to evaluate model performance.¹⁵⁹ The main consequences of the third and fourth choices (to use predictive modeling and experimental testing for model performance) are especially relevant here. Predictive models are “equations that have no obvious underlying physical or logical basis,”¹⁶⁰ and do not provide insight into the phenomena and people in the world the data points were harvested from. Such equations are highly sensitive to change, and may produce very different results based on any small or big difference in either data, analysis, or both. It takes a lot of additional modeling to come to some sort of reliable understanding of how an ML model will behave and how it can be used.¹⁶¹ This has important consequences for the type of questions that different outcomes from experimental testing trigger in the ML modelers that build them,¹⁶² for what type of understanding ML modelers seek, and what kind of ‘explainability’ they think a model needs to offer to subsequent users.

This confuses ‘explanation’ aims in places that systems are used in. For example, many explanation models that have been produced over the last years, and that are meant to represent what happens in ‘black box’ systems to support a practices’ transparency and/or accountability aims, are of more use to designers themselves.¹⁶³ ML designers are also in the best position to use such explanation models responsibly. In others, such explanation models may inspire an “illusion of engagement” between real world causes and model outcomes, as was discussed in the previous section. Since it is hard to explain the *true* type of understanding that these models afford, such ‘explainability’ design can become “a dangerous distraction, too narrow in its goals and fragile in

158 Gürses and van Hoboken, ‘Privacy after the Agile Turn (Version 3)’; Marijn Sax, ‘Optimization of What? For-Profit Health Apps as Manipulative Digital Environments’, *Ethics and Information Technology*, 3 January 202; Greenfield, *Radical Technologies*.

159 Malik, ‘A Hierarchy of Limitations in Machine Learning’, 2.

160 James Larus and others, ‘When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making’ (ACM 2018) 9 <<https://dl.acm.org/citation.cfm?id=3185595>> accessed 25 October 2018.

161 Malik, ‘A Hierarchy of Limitations in Machine Learning’; Larus et al, ‘When Computers Decide’.

162 Amore Cloud Ethics: Algorithms and the Attributes of Ourselves and Others 47–48.

163 Umang Bhatt et al, ‘Explainable machine learning in deployment’, in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, FAT* ’20 (Barcelona, Spain: Association for Computing Machinery, 2020), 648–57, <https://doi.org/10.1145/3351095.3375624>.

how it is understood.”¹⁶⁴ Meanwhile, behavioral insights about what types of reasoning decision subjects appreciate when they are given explanations are being promoted as useful input for AI explainability designers.¹⁶⁵ Such insights themselves don’t claim to understand what people need, but what they want, which makes some writers warn that machine-issued explanations should not be optimized “for lulling people into accepting an explanation that they can’t logically follow.”¹⁶⁶

Shallow claims about what people seek when they ask for explanations are also encouraged by long standing traditions of comparing people’s brains with computers.¹⁶⁷ This is where denials of called-out explanation crises enter. A typical argument is that algorithmic opacity is not as problematic as is claimed, if one corrects our base-line norms. This can be done based on cognitive, behavioral, and neurological research, or rather on the conclusions some tend to draw from it: that we have overrated human self- and other-understanding. In comparison to our ‘black box’ neurological functioning, AI systems should be regarded as no less opaque, and even *more* understandable. The kind of explanations that machines can generate about their conclusions “may already hit the mark.”¹⁶⁸ This line of reasoning was encountered at numerous moments in the course of thesis research. The (flawed) comparison saw much uptake in recent decades through the development of ‘artificial neural networks.’ These are literally compared to human brains. By now, neurologists deplore the popular association, concerned that it has narrowed the necessary scope and imagination of *medical* researchers, too.¹⁶⁹

Authors warn about the loss of knowledge of, and faith in, critical social human traditions of coming to mutual understanding,¹⁷⁰ and of holding each other to account. They critically engage with those who argue that correlations are a sufficiently reliable knowledge base,¹⁷¹ that experimenting with analytical weights in models do not constitute decisions with moral dimensions,¹⁷² and that politically sensitive inferences are knowable by obvious qualifiers which can then be adjusted, notwithstanding

164 Malik, ‘A Hierarchy of Limitations in Machine Learning’, 22–24.

165 Tim Miller, Piers Howe, and Liz Sonenberg, ‘Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences’, nr. arXiv:1712.00547 [cs.AI] (2 December 2017), <https://arxiv.org/abs/1712.00547> para.2.2.

166 Andrew D. Selbst and Solon Barocas, “The Intuitive Appeal of Explainable Machines (Draft),” *Fordham Law Review*, *Forthcoming*, February 19, 2018, <https://papers.ssrn.com/abstract=3126971>.

167 For a detailed history of all the forms this has taken, the types of communities that it thrives in, and critique, see: Siri Hustvedt, *A Woman Looking at Men Looking at Women: Essays on Art, Sex, and the Mind* (New York: Simon & Schuster, 2016) Part II: Illusions of Certainty.

168 Zerilli et al, ‘Transparency in Algorithmic and Human Decision-Making’.

169 Matthew Cobb, ‘Why Your Brain Is Not a Computer’, *The Guardian*, 27 February 2020, sec. Science, <https://www.theguardian.com/science/2020/feb/27/why-your-brain-is-not-a-computer-neuroscience-neural-networks-consciousness>.

170 Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, 168.

171 Amoores, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*.

172 Amoores, 163.

established research that says otherwise.¹⁷³ Still, the sketched developments do raise important questions to ask of our traditional explanation paradigms. When the bodies of knowledge that ground our decisional processes are driven by different natured pursuits of ‘what is important to know,’ assumptions about the functioning of our ‘good old fashioned’ explanation rules may be ill-advised and even risky. Consider a paper that investigates the depth of underlying knowledge that explainers need to have according to Administrative explanation rules. The authors conclude that it does not amount to a problematic benchmark with regard to many ADM systems.¹⁷⁴ But such a conclusion really depends on the depth of understanding of the Administrative explanation paradigm that was sought by the author. Consider that the types of harms that algorithmic systems are typically producing are not that novel, themselves. If existing explanation rules have not been very productive in understanding how this happens, we have no reasons to keep using them. The next sections take up this premise.

2.3 Problems of knowledge: four provocations of ‘new’

The previous sections discussed an array of perceived challenges to established aims of regulated explanation practices. The problems are seen to arise as a consequence of new methods and practices of knowledge making and decision making. This part of the chapter teases out an array of problematic premises that the presentation of these challenges as ‘new’ seem to require and presents these in the form of four provocations. Based on these provocations, the last part of the chapter will argue for an angle of research into the ‘right’ crisis: our weak understanding of what explanation rules should fundamentally aim to achieve for people, and whether they are designed to do so.

The analysis in these two parts already introduce some arguments from (mainly) philosophical domains of research that are known as (and relatable to) epistemic justice and -injustice. Authors in these domains investigate how methods and practices of knowledge making can amount to wrongful, themselves, as well as what can be done to avoid this. To quote Grasswick “[i]njustices deal in social relations and interactions. Epistemic injustices exist because a large portion of our epistemic lives are social.”¹⁷⁵ In doing so, the work provides valuable angles, and arguments, for an engagement with fundamentally (ir)responsible, obscure, and (in)humane explanation practices.

173 Balayn and Gürses, ‘Beyond-Debiasing: Regulating AI and its inequalities’; Bryson, ‘The origins of bias and the limits of transparency’; Synced, ‘Yann LeCun Quits Twitter Amid Acrimonious Exchanges on AI Bias’, *Synced* (blog), 1 July 2020, <https://syncedreview.com/2020/06/30/yann-lecun-quits-twitter-amid-acrimonious-exchanges-on-ai-bias/>.

174 Cary Coglianese and David Lehr, ‘Transparency and Algorithmic Governance’, SSRN Scholarly Paper (Rochester, NY, 9 November 2018), <https://papers.ssrn.com/abstract=3293008>.

175 Heidi Grasswick, ‘Epistemic Injustice in Science’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017).

But the fields are (as yet) rarely picked up in ‘right to explanation’ debates, and their insights need translation. The chapter after this one sources this work extensively and applies it more precisely.

2.3.1 Sense making on individual levels as a lucky draw

It was discussed how explainees are seen to be challenged in making useful sense of how decisions are made about them, because consequential choices for and about them are made in opaque parts of the system(s), and on non-individual levels. The increasing messiness and complexity of how this happens exacerbates the problem, and leads to doubt about the merits of proposed solutions such as those in data protection laws. The ‘enabling’ aims of these laws, such as to ‘meaningfully position’ explainees and reduce the effects of information inequality are questioned. Fundamental concerns are raised about respect for individual person-hood: autonomy, reasoning, and the right to understand one’s environment.

But investigations of historical, as well as contemporary choices in knowledge and decision-making methods show that making sense on individual levels has been traditionally hard, and made hard, for less powerful groups of individual explainees. Whole ‘populations’¹⁷⁶ were excluded from shared knowledge (making) spheres by colonizing powers on the grounds that they were less able to make sense, less worthy of engaging with intellectually, among more instrumental reasons.¹⁷⁷ And although in historical medical times most patients did not receive explanations, reasons for it differed: some groups of patients were considered to have no use for knowledge about their states because they did not have much power over their bodies in the first place.¹⁷⁸ Since then, medical research practices have ignored various groups of people to the extent that they still can’t get the explanations (and the treatment) that other groups can.¹⁷⁹ Investigations of such examples show the influence of wrongful ideologies on patterns of social and informational power abuse.¹⁸⁰ These patterns, and the groups, are similar to the wrongs that are revealed to happen through ADM. It has become established knowledge that ‘big data’ practices tend to have a disparate

176 cf Butler, the use of scare quotes means to say that these populations aren’t a ‘sociological given,’ rather, they are created by the very practices that treat them as such. Judith Butler, *The Force of Nonviolence: An Ethico-Political Bind* (Verso Books, 2021).

177 Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., ed., *The Routledge Handbook of Epistemic Injustice*, paperback edition (Routledge, 2017).

178 Jay Katz, *The Silent World of Doctor and Patient*, 1984, Johns Hopkins edition (Johns Hopkins University Press, 2002).

179 Jessica Nordell, ‘The Bias That Blinds: Why Some People Get Dangerously Different Medical Care’, *The Guardian*, 21 September 2021, sec. Science, <https://www.theguardian.com/science/2021/sep/21/bias-that-blinds-medical-research-treatment-race-gender-dangerous-disparity>.

180 The effects on groups who are treated this way are grave, as will be further discussed in the next chapter.

negative impact on groups of people that are already discriminated against, or who are marginalized in other ways.¹⁸¹

Previous sections discussed some arguments in favor of added ADM explanation rules that go beyond the typical justification of general rule-to-individual application. It was suggested that it may be necessary to explain what ADS are designed for, why they are employed in a context, under whose power they are developed, how they are governed and overseen, and to certify that knowledge and methods were investigated in terms of possible discrimination. There may be merits to such arguments, but the added explanation work they propose to do would have been necessary already. Historical studies of discriminatory and marginalizing policies already reveal a traditional intertwining of public funds, political will, behavioral and other scientific ‘progress’ and commercial technological lobby.¹⁸² It is also no secret that the required information is actively suppressed by companies and institutions who fund, govern, develop, use and sell the new knowledge resources.¹⁸³ Knowledge about how discrimination and marginalization ‘works’ technologically has not been amplified enough by authoritative voices in the AI field,¹⁸⁴ and neither are public institutions keen on admitting to, and sharing useful information about, racist and other institutional discriminatory dimensions of algorithmic ‘mishaps.’ The Dutch Benefits Scandal is a case in point.¹⁸⁵ To conclude; it is right that individual sense-making is tabled as a problem. But to present this as a new problem is historically ignorant and potentially subversive. When the reasons for our explanation problems are not properly identified, solutions risk to perpetuate the situation for those who need solutions the most.

181 Solon Barocas and Andrew D. Selbst, ‘Big Data’s Disparate Impact’, *California Law Review* 104, nr. 3 (2016): 671–732; Alexandra Chouldechova, ‘Fair prediction with disparate impact: A study of bias in recidivism prediction instruments’, *arXiv:1703.00056 [cs, stat]*, 28 February 2017, <http://arxiv.org/abs/1703.00056>; Cathy O’Neill, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016); Linnet Taylor, ‘What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally’, *Big Data & Society* 4, nr. 2 (1 December 2017).

182 Lepore, *If Then*; Benjamin, *Race After Technology*; Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*; Greenfield, *Radical Technologies*; Bridges, *The Poverty of Privacy Rights*; Black, *IBM and the Holocaust: The Strategic Alliance Between Nazi Germany and America’s Most Powerful Corporation*; McQuillan, *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*.

183 Reuters, ‘Google Told Scientists to Use “a Positive Tone” in AI Research, Documents Show’, *The Guardian*, 23 December 2020, sec. Technology, <https://www.theguardian.com/technology/2020/dec/23/google-scientists-research-ai-positive-tone>; Julia Powles and Hal Hodson, ‘Google DeepMind and Healthcare in an Age of Algorithms’, *Health and Technology* 7, nr. 4 (1 December 2017): 351–67.

184 Synced, ‘Yann LeCun Quits Twitter Amid Acrimonious Exchanges on AI Bias’.

185 ‘Xenophobic machines’; Achbab, ‘De Toeslagenaffaire is ontstaan uit institutioneel racisme’.

2.3.2 Idealistic assumptions about knowledgeable participation

It matters against which benchmarks for subject participation the above-described concerns about AI-infused practices are set off against. Idealistic assumptions about subject participation are common. Typical legal explanation paradigms assume that decision subjects can already anticipate decisional outcomes to some extent because they (sufficiently, if not entirely) understand the rules, rights, and procedures that are applicable to them, and that they are sufficiently knowledgeable to participate in the creation of information about them. All these abilities are argued to be under pressure. Data protection laws testify to what needs to be repaired: explainees need rights of access, correction and addition of information, of contestation, and to “meaningful information about the logic involved, as well as the significance and the envisaged consequences of [ADM] processing.”

But just as was true for the previous section, the challenges that are being called out as problematic are not so much new as they are traditional. When the right patterns are studied it turns out that the (types of) groups of people that were historically excluded from informed, responsible participation in decision making about them are not so different as the ones that are disproportionately affected now. E.g., for those with less fortunate positions in Dutch societies, *responsibly* navigating complex administrative landscapes and engaging in meaningful State interaction was already mostly illusionary. This was no secret. Scientific research reports about it were ignored by consecutive governments’ legislative and policy efforts.¹⁸⁶ And again, suppression of information is a pattern. Phone calls and correspondence of desperate citizens is routinely kept out of case files.¹⁸⁷ The medical domain deserves to be named again, too. Responsible participation was and is made impossible when certain groups of patients’ accounts of what ails them are routinely downplayed and ignored.¹⁸⁸ On a higher level, whole communities’ participation in public health schemes suffer because of understandable distrust, informed by (historical) accounts of maltreatment—a situation that, if not foreseen and addressed through honest communication and justification, simply

186 B.C. Filet, *Kortsluiting met de bureaucratie : over participatiemogelijkheden van burgers bij het openbaar bestuur*, Bestuur-Bestuurden, 1974; Claudia Kammer en Liza van Lonkhuyzen, ‘Oud-minister Bussemaker gelooft niet meer in de participatiemaatschappij’, *NRC Handelsblad*, 14 February 2019, <https://www.nrc.nl/nieuws/2019/02/14/misschien-was-ik-naief-a3654165>; Marieke Stellinga en Petra De Koning, ‘PvdA vindt eigen Participatiewet mislukt’, *NRC*, 11 November 2020, <https://www.nrc.nl/nieuws/2020/11/11/pvda-vindt-eigen-participatiewet-mislukt-a4019751>.

187 Esther Lammers, ‘Oud-ombudsman Alex Brenninkmeijer ziet in de toeslagenaffaire geen bedrijfsongeluk maar een falend systeem’, *Trouw*, 31 December 2020, sec. politiek, <https://www.trouw.nl/gs-bdc55fe5>.

188 Havi Carel en Ian James Kidd, ‘Epistemic Injustice in Medicine and Health Care’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017).

persists.¹⁸⁹ And in education, analog systems of ‘scoring’ have meant that ‘rewards and punishment’ have been dealt out based on circumstances and information that decision subjects have no influence on. Pupils are barred from (further) schooling based on their performance in tests that aren’t able to do them justice in the first place.¹⁹⁰

What seems to be *new* about the contemporary ADM-driven situations is the increased diversity of public that are prevented from responsible participation. It probably helps that the quality of correlative conclusions that are being drawn, and the actual effects of ML systems in practice have become harder to predict for those who design and implement them as well. The systems need to be experimented with, before and after implementation, which creates new risks for more people. The fact that preexisting problems have been allowed to ‘broaden and deepen’ to the extent that they have raises various questions about the aims of preexisting explanation regulation: for example about the envisioned measure and inclusiveness of explained ‘participatory ability.’ Especially in the European region that self-identifies as one where governance is driven by a high regard for individual autonomy and humane, dignified treatment.

2.3.3 Who respects whose human ability to reason?

The absence of causal relations in ADM ‘reasoning’ was named as a salient conceptual challenge with regard to standing requirements for the human explanation of decisions. ML Systems aren’t meant to sustain, find, or act on causal hypotheses. Section 2.2.4 suggested that there is a growing divide between those who value and make use of causal, and qualitative explanations, and those who are unconvinced that either of them are fundamentally necessary. The latter group is historically over-represented in the AI community, with effects on what kind of technologies are developed.¹⁹¹ It was discussed how especially ‘predictive’ systems are not just changing, and limiting possibilities for individual decision explanations, but are also causing confusion. Such systems produce statements based on past correlated instances: they conclude what

189 Rachel Humphreys, Nazia Parveen, and Annabel Sowemimo, ‘Vaccine Hesitancy: What Is behind the Fears Circulating in BAME Communities?’, *The Guardian - Today in Focus*, 26 January 2021, <https://www.theguardian.com/news/audio/2021/jan/26/vaccine-hesitancy-what-is-behind-the-fears-circulating-in-bame-communities-podcast>.

190 Ben Kotzee, ‘Education and Epistemic Injustice’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017); Macarthur R. S. and Elley W. B., ‘The reduction of socioeconomic bias in intelligence testing’, *British Journal of Educational Psychology* 33, nr. 2 (13 May 2011): 107–19; For a brief on the problematic history of IQ testing, see McQuillan, *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*, 88.

191 Broussard, *Artificial Unintelligence*; To add another social factor at play, global commercial-technological power (im)balances also express in structural technological design of knowledge and information systems, which can squeeze room and influence of different local, scientific, and cultural, practices of knowledge making and sharing that these technological structures are not conducive to Lilly Irani et al, ‘Postcolonial Computing: A Lens on Design and Development’, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’10 (New York, NY, USA: ACM, 2010), 1311–20, <https://doi.org/10.1145/1753326.1753522>.

could have been predicted to occur, but only in hindsight. They cannot predict what people will do, what will befall to them, and why this happens. Still such ‘predictions’ are used in precisely that way, to act on people and the world, which influences what befalls to people, and what they will be able to do.

But that this happens should not surprise anyone. The human inclination to search for, and act on, causal understanding is well established.¹⁹² The tendency to *ignore* such human understanding needs itself is what needs attention – and investigating this soon reveals that this should not surprise us either. Studying human behavior without an interest to understand what informs it, let alone to let the studied themselves define this, is a known and criticized research tradition.¹⁹³ It thrived in colonial and other wrongful ideology-driven practices, and has generally been popular in traditions that value quantitative over qualitative insight (which also tended to be exclusive with regard to who counted as practitioners.)¹⁹⁴ The tendency to be dismissive of the human ability to give accounts of themselves, and their reasons, were part of such traditions even before neurological blueprints were imagined to give the right kind of insight.

The historical tradition to value more quantitative, less qualitative ways of knowing over others thus stands in relation to idealizations of some groups of humans’ ability for ‘reason’ and ‘rational thought’ over others. The chapter’s introduction discussed how this should be kept in mind when assessing ‘dignitarian’ arguments for the fundamental right to explanation, which tend to foreground the individual human ability to reason. We won’t thrive equally when dominant methods of knowledge making exclude others and others types of understanding. Arguably, we won’t thrive at all, as the kinds of understanding that are excluded are simply fundamentally necessary. The discussed explanation challenges testify to that. Cited for example was how it turned out to be hard for ML developers to responsibly explain what a system’s ‘predictions’ actually were, even to highly educated audiences, which would include projected users of such systems. Keeping in mind how explanation moments are also knowledge making moments, the illustration showed how explainees came away with knowledge, but not necessarily the knowledge that the explainer meant to convey, which was that the correlative predictions of the system were *not* causal. The question to ask of the example is, whose lack of understanding does the situation testify to? When those who currently develop and explain technology tend to exclude ‘non-technological,’ especially qualitative insights about humans and the world, and the effects of their systems on and in them from their own understanding, their understanding is not sufficient for the kind of explanation others seek.

192 Helena Matute et al, ‘Illusions of causality: how they bias our everyday thinking and how they could be reduced’, *Frontiers in Psychology* 6 (2 July 2015): 888.

193 ‘Anthropology, for example, is a field whose origins were premised on the subordination of certain peoples.’ Grasswick, ‘Epistemic Injustice in Science’, 320; Lepore, *If Then*; Siri Hustvedt, ‘The Delusions of Certainty’, in *A Woman Looking at Men Looking at Women: Essays on Art, Sex, and the Mind* (New York: Simon & Schuster, 2016).

194 Hustvedt, ‘The Delusions of Certainty’.

Taking this further, such practices of development will also produce unhelpful solutions to algorithmic harms. Indeed, comprehensive critiques of proposed technological solutions are being published at the time of writing. Among other things these pertain to illusions of ‘debiasing’;¹⁹⁵ presentations of harms as the result of “errors, accidents, or aberrations” rather than of methodological choices,¹⁹⁶ and to how the focus on how technological solutions take necessary attention away from understanding how harms arise from types of knowledge making more broadly.¹⁹⁷ To cite Gebru, “It’s like ‘let’s diversify our data sets. And that’s kind of ethics and fairness, right?’ But you can’t ignore social and structural problems.”¹⁹⁸

The point these writers make is not that technological interventions can’t help to improve our technological systems, and useful strategies are certainly being developed.¹⁹⁹ The point, rather, is that this work needs to be part of a more comprehensive effort. This broader engagement also requires other methods of investigation, and of development, with regard to diagnosing how wrongs occur and how to deal with this. Epistemic injustice literature makes insightful how the process of dismantling unjust knowledge practices, which many algorithmic wrongs arguably are, needs analysis through “political, ethical, and epistemological philosophical endeavors.”²⁰⁰ Applied to explanations, the above considerations also advise to take a comprehensive look at solutions with regard to explanation regulation. Simply adding a layer of ADM-requirements may irresponsibly ignore a structural lack of protections in our current legal systems. What have we typically asked explainers to understand, and to justify, to their explainees?

2.3.4 Explainers are always challenged, whether they know it or not

This last provocation zooms in on the person of the explainer, and raises some questions about the challenges they are said to be facing when ADM systems are a factor in decisions that they are obliged to explain. Among these challenges were concerns that

195 Balayn and Gürses, ‘Beyond-Debiasing: Regulating AI and its inequalities’; Bryson, ‘The origins of bias and the limits of transparency’.

196 Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, 119; Malik, ‘A Hierarchy of Limitations in Machine Learning’, 1–2.

197 Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*; Malik cites Issa Kohler Hausmann in a critique on ‘counterfactual race’ efforts: taking race out of an equation to determine if a person has been discriminated ignores that ‘race’ is a factor in producing all other attributes that are counted. Malik, ‘A Hierarchy of Limitations in Machine Learning’.

198 Timnit Gebru and Emily Denton, ‘Tutorial on Fairness Accountability Transparency and Ethics in Computer Vision’, 2020, <https://sites.google.com/view/fatecv-tutorial/home>.

199 Timnit Gebru et al, ‘Datasheets for Datasets’, *ArXiv:1803.09010 [Cs]*, 23 March 2018, <http://arxiv.org/abs/1803.09010>; Margaret Mitchell et al, ‘Model Cards for Model Reporting’, *ArXiv:1810.03993 [Cs]*, 5 October 2018, <http://arxiv.org/abs/1810.03993>.

200 Gaile Pohlhaus, Jr., ‘Varieties of Epistemic Injustice’ in Ian James Kidd, José Medina and Gaile Pohlhaus, Jr. (eds), *The Routledge Handbook of Epistemic Injustice* (paperback edition, Routledge 2017).

decision makers/explainers strongly engaged in digital data-driven methods may be tempted to objectify decision subjects to their data aggregates. Another concern was their lack of technological understanding of the knowledge and methods they work with, and of how their own reasoning is influenced by the systems they work with. Explainers have become explainees, themselves. Still their roles as human moral reasoners and guardians of the enactment of individual justice is emphasized in regulatory solutions. But what, if anything, can they usefully justify when good-old-fashioned-rule to individual application is not how decisions are made anymore, when what needs to be explained may need to encompass how “models impose the logic of models on the world”?²⁰¹

To understand these challenges better, and respond to them, a first starting point for thought is an obvious one: all explainers are always inevitably only partially informed. How do we normally deal with this? Think of how physicians necessarily make use of other types of expertise, made by others in their field. (Including, as it happens, much quantitative and correlative knowledge.) They are also subject to larger power structures and their influences, e.g. public health policy and insurers, the interests of ‘big pharma’ and so on. As such, they must already face explanation challenges, and they already need to avoid to become instruments of ‘bad knowledge practices.’ As Amore writes, “[t]he apparent opacity and illegibility of the algorithm should not pose an entirely new problem for human ethics, for the difficulty of locating clear-sighted action was already present.”²⁰²

For human *law*, this is not so different. Think of the judges in the Childcare Benefits scandal. They were not made aware of the fact that algorithmic systems played a role in the decisions that victims appealed in their courtrooms. But this was hardly their biggest challenge in ‘locating clear-sighted, justifiable action.’ In their reflections, the Judges reported how they wrestled with fundamental legal principles, building blocks of the logic of judicial decisions: to treat equal cases equally, and to apply rules in similar ways across local jurisdictions to foster a knowable legal response.²⁰³ These principles inhibited them to reason ‘against the grain’ in individual cases.²⁰⁴ It also bothered them that the problems seemed to originate on other than individual levels. One judge reflected that the merits of claiming individual injustices should perhaps have been outweighed by the merits of claiming *systemic* injustice, and presenting such claims to the Administrative Supreme Court who routinely struck rulings that favoured explainees, siding with the State.²⁰⁵ The accumulation of the many individual cases

201 Malik, ‘A Hierarchy of Limitations in Machine Learning’, 13.

202 Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, 8.

203 Principles of equal treatment, foreseeability and legal certainty.

204 ‘Lessen uit de kinderopvangtoeslagzaken. Reflectierapport van de Afdeling Bestuursrechtspraak van de Raad van State’, 8.

205 ‘Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken’, 43.

of injustice into what broadly became acknowledged as “a failure of constitutional democracy”²⁰⁶ was foreseeable, and was ignored.²⁰⁷

The example illustrates how explainers have an inevitable role in establishing, sustaining, proliferating, but also in resisting, and countering, knowledge and decision methods and practices. And how their individual capacities and capabilities to fulfill this role should not be exaggerated or idealized. This easily becomes as misleading as the attribution of too much wrongful and ameliorating powers to technology. There is also the risk that human explainers are sidelined as useless, which would fit neatly into the arguments of those who downplay human reasoning as more ‘black box’ than that of ML systems. And that is a very bad idea.

As an example of the need for a more comprehensive approach, in the Netherlands, senior elementary school pupils are scored for admittance to any of 5 levels of Dutch secondary (‘high school’) education.²⁰⁸ The score has alternatively been based on standardized testing, teacher’s assessments, and combinations thereof. There are ‘forever’ discussions about the merits of either choice, and concerns about discriminatory biases in the eventual scores are justified. But although research has repeatedly shown how several institutional dynamics produce these patterns, primary school teachers’ biases tend to attract the most attention.²⁰⁹ Researchers are particularly concerned that this leaves important drivers unaddressed and idealizes standardization tools that reproduce societal biases. E.g, the standardized tests are better aligned with the output of privileged children—a common problem in other parts of the world too.²¹⁰ And affluent parents invest in additional training for their children’s standardized tests.²¹¹ In a country where affluence is correlated with ethnicity, this is a factor of importance.

206 Judge Sprakel, ‘Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken’.

207 Judge Cooijmans, ‘Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken’.

208 Upon successful completion of a lower level of secondary education, depending on the courses they took, pupils can get access to higher level education. But stacking (‘stapelen’), as its called, has become increasingly hard. This is not further discussed here.

209 L Borghans en R.E.M. Dieris, ‘Ongelijkheid in het nederlandse onderwijs door de jaren heen’, in *Preadviezen voor de Koninklijke Vereniging voor Staathuishoudkunde*, edited by A Gielen, D. Webbink, en B. ter Weel (ESB & Koninklijke Vereniging voor Staathuishoudkunde, 2021).

210 Kotzee, ‘Education and Epistemic Injustice’; Macarthur R. S. and Elley W. B., ‘The reduction of socioeconomic bias in intelligence testing’.

211 ‘Trainen voor de test’, *De Groene Amsterdammer*, April 2014, <https://www.groene.nl/artikel/trainen-voor-de-test>.

The example of educators is important: educators are explainers *par exemple*. And they can certainly be biased, just like judges can be. So far, corrective algorithmic solutions have not made things better, sometimes worse, and understanding how this happens adds investigative burdens.²¹²

*

To conclude, this section's provocations argued that concerns about the loss of individual subject understanding and participation, about objectifying group treatment, deferred and demoralized explainer responsibility, and about the challenges of explainers to make the quality of decisions insightful are well-placed, but when presented as 'new,' also dangerously distracting. Not just in light of preexisting instances of such problems, but in light of how our technological methods were in the making for decades, driven by thoughts and ambitions whose ontologies have a much longer history.

Explanations in the domains and situations the thesis focuses on are assumed to play an important role in protecting decision subjects from the abuse of power by actors who make consequential decisions about them. These sections raised several questions about these idealistic expectations, especially in light of the pervasive existence of problematic methods and practices of knowledge making. It also considered how explanations are knowledge making practices themselves. The next sections continue this line of thoughts, and settle on a line of questioning to pursue in the written thesis.

2.4 Are our explanation rules in need of justification?

2.4.1 Two meaningful applications of 'meaningful'

2.4.1.1 Meaningful information positions

In research and governance discourse on solutions to our (perceived) explanation challenges, a much-used term is 'meaningful.' Meaningful information, meaningful oversight, meaningful human intervention for example features in Authoritative guidelines on the GDPR's explanation regime.²¹³ The term's intuitive appeal is broadly shared, but its elaborations are seen wanting. E.g., appeals for meaningful oversight

²¹² Jeff Larson, Surya Mattu, Lauren Kirchner and Julia Angwin, 'How We Analyzed the COMPAS Recidivism Algorithm', ProPublica, last consulted 15 December 2020, <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=XSO7CCiM7D0udJrFYQeZnvAitR3ZT0sj>; Reuben Binns, 'Human Judgment in Algorithmic Loops: Individual Justice and Automated Decision-Making', *Regulation & Governance* 16, nr. 1 (2022): 197–211.

²¹³ Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679 (wp260rev.01)'; Article 29 Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP 251rev.1)'.

remain under-defined,²¹⁴ or are criticized for how they sketch a simplified picture of our socio-technical worlds.²¹⁵ As was discussed, the highest EU court has been asked for an authoritative opinion on what ‘meaningful information about the logic’ should mean. In general, there seems to be some agreement that the whole system needs to be taken into account to seek out what it meaningful about it, and that in the eventual individual situations that the law sees to, what is meaningful will be highly contextual.²¹⁶ This puts decision domain experts in the spotlight, including a domain’s designated explainers. The term meaningful is indeed applied to them, too: they are the ones that are relied on to meaningfully intervene, provide meaningful information, and in general infuse the process with meaningful, *humane* interaction. But these uses of meaningful, too, do not explicate what it is they should be doing, or re-instating. A usable clue can be found in a much-cited paper by Binns et al, which explores how analog explainee perceptions of justice are triggered in ADM contexts. They described how test subjects of *machine*-issued explanations seemed to seek an equivalent of ‘interactional justice,’ described as “being treated with dignity respect by the decision-makers.”²¹⁷ This suggests, that there is a social relationship between explainer and explainee that also expresses itself in the information that is given, and the authors suggest this as a subject for further research. However, in light of the provocations in the previous sections, perhaps that challenge should be taken up regardless of how technological the context is. Through yet another example from the Childcare Benefits Scandal, this section introduces the merits of defining ‘meaningful information positions,’ or, as a verb, ‘meaningful information positioning,’ as an expression of interactional justice, and suggests what needs to be further investigated to operationalize this notion.

Judges from the court of first instance reported how up until and including sentencing, they had been unaware of the involvement and methods of the Tax Administration’s fraud teams in decisions about their complainants. But much other information that the Tax Administration could or should have provided was lacking in case files, too, and this *was* apparent. Crucial documents were also routinely added immediately before trials, too late to study. In other words, the information positions of judges were (very) badly served. Looking back, some of them wondered why they had not made more of a fuss about this. The messiness turned out to be a clue: it fit a pattern of obfuscation that also included critical information about the fraud teams, an omission that turned

214 Ben Green, ‘The Flaws of Policies Requiring Human Oversight of Government Algorithms’, *Computer Law & Security Review* 45 (2022): 16–17.

215 Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*; Green, ‘The Flaws of Policies Requiring Human Oversight of Government Algorithms’.

216 Edwards and Veale, ‘Slave to the Algorithm?’; Selbst and Powles, ‘Meaningful Information and the Right to Explanation’; Asghari, Hadi et al, ‘What to Explain When Explaining Is Difficult. An Interdisciplinary Primer on XAI and Meaningful Information in Automated Decision-Making’ (Zenodo, 22 March 2022), <https://doi.org/10.5281/ZENODO.6375784>.

217 Reuben Binns et al, “‘It’s Reducing a Human Being to a Percentage’”; Perceptions of Justice in Algorithmic Decisions’, in *Proceedings of the 2018CHI Conference on Human Factors in Computing Systems*, 2018, 1–14.

out to be unlawful in light of the fraud team's crucial roles.²¹⁸ The report based on their reflections also advises that they serve their own as well as their explainees' information positions better: to investigate situations well enough to be able to appraise the facts of the cases; to pursue more understanding about the positions of explainees in the larger decisional space that they were subject to.²¹⁹ Victims themselves had been kept in the blind by the Tax Administration throughout their ordeals, in other words, they were kept from participating meaningfully and in their own interest. And instead of a well-informed judge, they were met with (self-expressed) ignorance and disinterest. The explanation in the form of a verdict that found nothing 'legally unacceptable' in what had happened to them arguably added insult to injury.

The explainers in this example are domain experts in their field, and relied on for meaningful human review, intervention, interaction and explanation. Their reflections—above, and those cited in earlier sections—table several things that (should) make their *information positions meaningful*. They expect themselves to critically engage with rules and fundamental principles of their domain; to use their moral conscience and intuition to signal injustice; to pursue clues that they might be missing information; to ensure that they know about and understand decisional aims, methods and processes of decisions under scrutiny; to qualify the informational positions of their explainees; and to understand the effects of their judgments in the larger societal (administrative, in this case) system their explainees are subject to. But an additional suggestion follows from the previous sections' discussions. For example, even if these judges would have been (made) aware of the existence, aims, and methods of the Tax Administration's fraud teams, their investigative jobs would have arguably continued. Because, although the Tax Administration's 'heartless' behavior and ruthless aims have been broadly condemned, at the time of the reflection report, no legal, political and societal agreement existed on whether the engaged in 'ethnic profiling,' even where nationality was not used²²⁰ nor on what other instances of (institutionalized) racism expressed in their behavior, and in their algorithmic systems.²²¹ These questions remained judicially uninvestigated, and unaddressed in judgments (explanations) even though the apparent 'ethnicity' of most of their explainees begged the question, and even though such patterns in Dutch administrative decision making are known to exist.²²² This arguably reduced the quality of the explanatory exchanges as well as the outcomes, which did not contribute to the information positions of the scandal's victims. It also raises questions about the judges' understandings of 'justice' to be on the lookout for. Arguably, the omission means that the racism involved (as has been acknowledged at the time of thesis writing) was *reasoned away*.

218 'Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken', para.4.7.

219 'Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken', section 5.1.1 and throughout.

220 Merel Koning, Amnesty: the Autoriteit Persoonsgegevens uses a restricted definition of 'ethnic' 'Xenophobic machines'; Besluit tot boeteoplegging Minister van Financien.

221 Achbab, 'De Toeslagenaffaire is ontstaan uit institutioneel racisme'.

222 Çankaya, 'Opinie | Ze bedoelden het wél zo – het racisme kan onmogelijk ontkend worden'.

In all the reports and discussions on the scandal, not much attention was paid to the role of explanation *rules* as of instructive influence on what makes decisions reasonable: those of the judges, and of the civil servants in the chain of decision making that led up to the court cases. But one has to wonder.

2.4.1.2 *Meaningful explanation rules?*

The start of the chapter introduced the practical phenomenon of explanation rules: what they are, and what they—generally—mean to do. This section introduces explanation rules again, but this time in a more speculative way. It discusses explanation duties as instructive statements about the pursuit of (both parties’) ‘meaningful information positions.’ The suggestion is sustained by arguments from the epistemic justice/injustice domains, with which these are further introduced.

The discussion of dignitarian explanation aims warned to not assume that ‘we thrive because we reason,’ even if reasoning lets us thrive. Also cited were critiques of understandings of dignity that are rooted in ideals of individual autonomy rather than humanity’s social inter-dependency.²²³ These concerns can inform the pursuit of meaningful information positioning as an expression of interactional justice. History shows how humanity’s fundamentally shared need for reliable knowledge to act on, and our ability to thrive when we have such knowledge, does not naturally lead to institutional social arrangements for the production, communication (explanation) and use of such knowledge; these are deliberate efforts.²²⁴ And so norms have been developed for e.g., scientific practice, for journalism, for medical research on humans. Norms are expressed in professional, ethical, and legal rules. But these haven’t prevented that participants of such practices also create *unreliable*, and harmful, rather than beneficial knowledge.²²⁵ Nor have they prevented that some groups are systematically excluded from participating in production.²²⁶ Setting norms for knowledge practices therefore matters greatly. And studying the norms we set, and how we set them, gives insight into the workings of fundamental, human, epistemic interdependence.

The explanation rules this thesis is concerned with are one such type of ‘institutional social arrangement.’ They contain norms for what needs to be explained about the purpose, process and outcome of decisions (used here in the broad sense—including consequential expert opinions, and recommendations.) Explanation rules are therewith a statement about what is of interest to understand, for whom; what a meaningful

223 Law, especially medical law ‘is at its very worst when it is motivated purely by a desire to placate the autonomous man,’ Foster, *Human Dignity in Bioethics and Law*.

224 Code, *Epistemic Responsibility*.

225 Harms in and of knowledge practices are discussed at length in the next chapter, the reader is referred to there.

226 Kristie Dotson, ‘Conceptualizing Epistemic Oppression’, *Social Epistemology* 28, nr. 2 (3 April 2014): 115–38.

information position is for them to start from, and to arrive at. This also pertains to what, if anything, should be understood and explained about knowledge and methods that underlie, and indirectly inform decision making. Even when they don't express this explicitly, explanation rules *inevitably* set standards for the understand-ability, and explain-ability of underlying knowledge. In addition, explanation rules inevitably assume that explainers and explainees can meaningfully understand each other. But explainers and explainees don't share all of their knowledge spheres, and they start from their own, more or less apparent, information positions. It would make sense for explanation rules to address these aspects, too. Do they?

As social arrangements go, our explanation rules are in a state of perpetual flux. E.g., in health care, the paradigm shift from 'doctor knows best' to informed consent rules entailed no less than a complete re-making of assumptions about what patients could, would, and should understand. And as will be discussed in Chapter 5, the shift unsettled explainers' profound assumptions about their knowledge domains.²²⁷ Such changes in our "govern-mentality"²²⁸ can be understood as progress. Ideally, the norms we set with regard to knowledge practices, including explanation, are alive to progressive developments in a societies' knowledge communities.²²⁹ That however raises some questions about our current explanation rules. Much recent algorithmic harm was predictable, and even predicted, but went undetected and unexplained in well regulated, arguably fundamental explanation domains: health care, Public Administration, the Judiciary. How progressive have we been?

To understand this better, existing explanation paradigms can be investigated. 'Good old' explanation rules indeed are being studied in light of the ADM explanation concerns, but arguably not yet in terms of the thesis's proposed notion of meaningfulness. They are mainly assessed to see whether and how they could deal with the challenges of *new* decision technologies. Different conclusions have been drawn. With regard to the legal domain, Some found that the bulk of explanation demands weren't all too complex, and AI could already comply.²³⁰ Others argue that the depths of legal explanation rules and principles still have more to offer and should guide further XAI research.²³¹ Useful warnings for conflating legal and technological concepts are voiced, such as 'probability' in criminal law, as this confuses our normative thought.²³²

227 A painful process, as famously described by Katz, *The Silent World of Doctor and Patient*.1984.

228 Bacchi Carol Bacchi, 'Why Study Problematisations? Making Politics Visible', *Open Journal of Political Science* 2, nr. 1 (26 April 2012). 5.

229 Anderson, 'Epistemic Justice as a Virtue of Social Institutions'; The legal codification of norms renders them objects of deliberate societal negotiations Hirsch Ballin, *Advanced Introduction to Legal Research Methods*, 23,103.

230 Cary Coglianese and David Lehr, 'Transparency and Algorithmic Governance', *Administrative Law Review* 71, nr. P.1 (2019).

231 Oswald, 'Algorithm-assisted decision-making in the public sector'.

232 Kiel Brennan-Marquez, 'Plausible Cause: Explanatory Standards in the Age of Powerful Machines', *Vanderbilt Law Review* 70 (2017): 1249–1301.

Cited earlier were authors who warned that when the knowledge making that informs decisions is effectively, but not officially, outsourced, the right to be believed as expert explainer should stop right there.²³³

This thesis undertakes a different level of study, one that investigates explanation rules' expression of 'interactional justice' aims, understood as meaningful information positioning, ultimately in service of the prevention of knowledge-induced wrongs.

2.4.2 Research question and sub questions

The previous sections suggested several questions to ask, several paths of inquiry to undertake to come to a better understanding of our contemporary 'explanation crisis,' and inform our explanation practices moving forward: to inform the practices we have, and those we will have in the future. It zoomed in on the duty of human explainers who serve the rights of explainees. Whether they mean to or not, they perform a function in the (further) production or prevention of 'harm by knowledge,' and are relied on to do the right thing. One would expect explanation rules to serve this function. The proposed inquiry requires further investigation of several 'stack-able' relations: between knowledge practices and the well-being of people, between this quality of knowledge practices and the responsibilities of explainers, and between such an understanding of explanation and the role of explanation rules. These questions are assembled into the following research question:

"In light of the existence of wrongful²³⁴ knowledge practices; understanding the explanation of decisions to decision subjects as 'knowledge making about knowledge making,' and understanding rules that govern explanation as regulating 'conduct about conduct,' how can explanation rules promote responsible explainer behavior?"

The breakdown of this question produces four sub questions:

Sub question 1 How do knowledge practices produce and avoid to produce harms?

Sub question 2 How can explainers avoid to re/produce epistemic harms in their practices?

Sub question 3 How do existing legal rules in two seminal regulated explanation domains promote responsible (non-oppressive, information position improving) explainer behavior?

233 Brennan-Marquez; Coglianese and Lehr, 'Transparency and Algorithmic Governance', 2019; Oswald, 'Algorithm-assisted decision-making in the public sector'.

234 cf. Veli Mitova, 'A New Argument for the Non-Instrumental Value of Truth', *Erkenntnis* 2021 (12 June 2021): 12.

Sub question 4 What lessons from the analysis of existing explanation regulation can we draw to inform how we deal with ADM explanation regulation?

Chapter 3 (the next chapter) takes on questions 1 and 2. Chapters 4 and 5 investigate the main explanation rules of the General Administrative Law act ('Awb') and the main explanation rules of the Medical Treatment Agreement Act ('WGBO') in relation to General (health) Practice.

2.5 Chapter 2 in a nutshell

This chapter discussed how explanation rules that see to decisional powers and powers of expertise mean to allow decision subjects and others to assess whether the treatment of a subject agrees with a society's rules and (other) norms. They are societal expressions of what is of interest to know about a decisional process, and what counts as sufficiently reasoned.

It then discussed various perceptions of contemporary challenges to the enjoyment of explanation rights, and the performance of explanation duties. It took this novelty framing to task by arguing that cited problems of individual subject understanding and participation, objectifying group treatment, deferred and demoralized explainer responsibility, and the undervaluing of inter human understanding were already suffered by less privileged group on a broad scale—and that exacerbating effects of automation were to be expected even if the kind of inscrutability that explainers now face is also different.

The chapter arrived at several decisions with regard to the delineation of the thesis's problem space. It argued to focus on a particular actor: explainers, who are made important to the preservation of fundamental explanation values in [novel] regulatory instruments. The chapter also argued to focus on a particular theme with regard to explainers' roles and regulatory duties: the existence of oppression in knowledge and its practices. It suggested to treat explanation *as* a knowledge practice: as 'knowledge making about knowledge making,' where the latter pertains both to knowledge that informs inform decisions (law, policy, expertise), and to decisional methods. Explanation rules, in such a view, govern 'conduct about conduct.' This angle informed the research questions and the (further) wisdom to seek from the research fields of epistemic in/justice.

Care to explain?

3 **Meaningful information positioning: an application of epistemic injustice and justice theory to explanation regulation**

3.1 **Re-idealizing explanation in recognition of non-ideal theory**

3.1.1 Arguments for a ‘negatively informed’ theoretical engagement

The use of increasingly inscrutable computational methods in decision making challenges established paradigms of regulated explanation. Fundamental concerns about fundamental explanation rights, aims, and duties are raised. When we don’t understand how and why things happen to us, we cannot hold each other to account. When we no longer understand our ‘environs,’ we cease to be able to navigate them safely. In these concerns, the ability to understand the knowledge that informs decisions, and that is constituted through decisional processes, is a precondition for the preservation of human freedom as protection from oppression. But who are the ‘we’ in peril? The previous chapter considered how the explanation challenges already existed, but not for everyone equally. Novel explanation challenges were predicted as well, but not for everyone equally. This raises questions about our existing explanation paradigms and their propensity for protecting explainees equally. If the fundamental needs of those who are less served are reasoned away under the same flag of dignity that is now waved so hard, calling on such fundamentals won’t serve humanity.

To perform the necessary investigation of our explanation paradigms, and to support contemporary efforts to amend, add to, or re-explain our explanation rules, this chapter models a set of technology agnostic ‘duties of care’ that should govern any practice of individual explanation of consequential decisions. This effort is aimed to support, and if necessary, to force the societal progress that our explanation rules are ideally alive to. But a specific engagement with the concept ‘ideally’ that was just used so casually is a prerequisite. The aliveness of explanation rules to societal developments needs to be an aliveness to all members of the knowledge communities the rules are meant to serve. And since there is reason to think groups tend to be excluded, we need to understand how such exclusion happens before we can responsibly imagine what proper explanation governance is.²³⁵ More precisely, a ‘re-idealization’ of explanation itself needs to be informed by insight into wrongful knowledge and explanation practices.

235 This ambition is co-inspired by Ruha Benjamin’s chapter “Retooling Solidarity, Reimagining Justice.” The re-imaginings of critical race theory cited by Benjamin remain necessary, and it is with humbleness that this thesis hopes to contribute to bring these insights forward into a domain where they are much lacking. Benjamin, *Race After Technology*, Ch. 5 Ruha Benjamin cites Derrick Bell on re-imaginings: “[t]o see things as they really are, you must imagine them for what they might be.”

This phrasing builds on Charles Mills's argument for the 'non-idealizing approach' to ethical or moral theory.²³⁶ Mills discusses several ways in which ethical idealizations, which in our case would be explanation justice, are inept to deal with the non-ideal (with explanation *in*justice) when they aren't sufficiently informed by how things actually go wrong for groups of people. Notably, in terms of their institutional oppression, subordination (whether on the basis of class, ethnicity, gender, geo-location..) This is an especially grave choice for theories that pretend to be universally beneficial. In our case, the notions under scrutiny are fundamental explanation needs and values, such as the dignitarian arguments, including the intrinsic value of explanation. Several of Mills's points resonate clearly and usefully with concerns about the theoretical underpinning of our ruled explanation paradigms, and/or how these work out for explainees. The rules that are the focus of the thesis are legal, not ethical. But they "create normative expectations about the behavior that the rules are intended to guide," and as such express normative ideals about that behavior. For the purpose of 're-idealization,' these normative ideals are the object. The legal-ethical difference will be made to matter again in the last section of this chapter that argues that the re-idealized notions need to have a place in the legal paradigm.

The first point Mills makes is that ideal theories have tended to start from 'atomistic' individuals as classic liberal ideal-types.²³⁷ The previous chapter introduced critique on this ideal already, as well as more inclusive (and realistic) views on autonomy, and dignity, as 'relational.' But Mills' argument goes further. The individualistic ideal abstracts away from relations of structural oppression, and so, it does not account for how the idealization embeds these power inequalities.²³⁸ In addition, not engaging with –historical and contemporary– oppression and its legacies obscures the shaping influence of oppression on our "basic social institutions (as well as the humans in those institutions)."²³⁹ Therewith, these are unhelpfully idealized too. In our case, this would pertain to our basic, institutionalized explanation practices such as those in law and medicine, and the explainers in them. Most importantly for the thesis's objective, Mills warns for idealized "cognitive spheres." Such idealizations assume shared knowledge and mutual understanding where in fact, the "social cognition" of participants of unequal powers will have been influenced by their different group experiences.²⁴⁰ When these kinds of idealizations are at play, obstacles to understanding that are presented as challenges to individual autonomy, or as 'intrinsic' difficulties in understanding the world, are problematizations that probably won't help to improve the situation for all

236 A strategy 'best and most self-consciously' developed by Onora O'Neill; as cited by Mills: "the best way of realizing the ideal is through the recognition of the importance of theorizing the nonideal" Charles W. Mills, "'Ideal Theory' as Ideology", *Hyppatia* 20, nr. 3 (2005): 166.

237 Mills, 168.

238 This can also apply to the attribution of idealized capacities that abstracts from how the opportunity to develop these have fallen to a small privileged group. Mills, 168.

239 Mills, 169.

240 Mills, 169.

‘epistemic agents’ equally. This gives an additional argument to ‘distrust’ the similar presentation of problems that chapter one discussed, which already deserved scrutiny for how they were presented as ‘new.’ To cite Gordon, “[o]ne of the signs of bankrupt positions is the need for them to hide under the guise of the new.” The remark is made in the context of the argument against (especially) historical philosophical efforts of the Global North to ‘hoard reason’ and exclude epistemic contributions from the Global South.²⁴¹ This has led to “distorted reason”: unreasonable and closed, rather than open and relational.²⁴² When such reason is used to reason away the rights of others, “reason is in need of justification.”²⁴³

These points will be returned to in the sections that follow. Mills is one of the consulted writers on notions and matters of epistemic injustice, the ‘non ideal theories’ this thesis chooses to engage with for its re-idealization of explanation. Such injustices refer to “any relation that disadvantages someone as an epistemic agent,”²⁴⁴ the someone in our case being the explainee. Notions of epistemic justice, contrarily, describe how to avoid this.

3.1.1.1 Note on literature

As explained in the Introduction chapter’s methods section, the acquaintance with the fields of Epistemic In/Justice was a gradual process. The learning that started in the course of the thesis project became a lasting engagement that will keep developing over time. The notions became a productive ‘voice over,’ resonating with themes of interest that were already present and informing current work and future plans. This however also makes it hard to ‘close’ this thesis’s early engagement and report on it without falling into a trap of reflective iterations.

To add to what was said about the approach of these bodies of work as described in the methods section, this section (only) adds some words on books that were used as major sources. On the ‘justice’ side, Feminist epistemology and philosophy of science scholar Lorraine Code’s 1987 *Epistemic Responsibility*²⁴⁵ analyses humanity’s inherent sociality and fundamental interdependence with regard to knowledge practices. She argues and describes the moral responsibility this puts on the shoulders of all, but especially on those ‘officially’ in the business of knowledge making. Within this group she emphasizes the responsibility of those who form ideas *about* knowledge making: philosophers, especially those concerned with epistemology. She is critical of how they tend to either

241 Gordon, *Freedom, Justice, and Decolonization*.

242 Interview with Madina Tlotsanova, Gordon, 128, Epilogue.

243 Lewis Gordon, ‘Shifting the Geography of Reason’, <https://www.uva.nl/en/shared-content/faculteiten/en/faculteit-der-geesteswetenschappen/events/events/2022/05/spinoza-1.html> Printed edition forthcoming, AUP.

244 Mitova, ‘A New Argument for the Non-Instrumental Value of Truth’, 9.

245 Code, *Epistemic Responsibility*.

abstract from human sociality, or defer to it with relativist arguments. She cites Bernard Williams, but he does not cite her, even though their ideas align on important points and hers were published earlier.

Bernard Williams's 2002 book *Truth and Truthfulness*²⁴⁶ focuses on the fundamental values of truth-finding. These values describe a 'proper' social epistemic practice that all of humanity fundamentally relies on, and that is especially important to prevent knowledge-related oppression in unequal social power relationships. Williams presses the moral importance of maintaining the kind of epistemic faculties, to nurture the kind of disposition that allow to engage with social powers at play in human knowledge practices responsibly. Although Williams does not (much) use the term justice, the core elements of his 'truthfulness' are aligned with descriptions of epistemic justice that are described by other authors. A weak point in this work is Williams's reliance on societal norms and traditions (such as those of 'research') without giving too much thought to how unstable these are.

Machteld Geuskens' 2018 doctoral thesis *Epistemic Justice: a Principled Approach to Knowledge Generation and Distribution* sustained these early engagements and helped to build a bridge to law and regulated environments. Her focus on what responsible sharing of knowledge means in light of what characterizes knowledge as a concept was the first clear bridge to practices of responsible explanation, and her mapping of related fields and authors further helped to find my own bearings. The thesis's more explicit engagement with notions of epistemic *injustice* also provided the right kinds of leads out to the works of authors from this closely related field.

While sourcing the research field of epistemic *injustice* for articles relevant to this thesis's subject space, several authors stood out for their comprehensive introductions of the field's ontology and development. Many of these authors joined hands to write *The Routledge Handbook of Epistemic Injustice*²⁴⁷ (hereafter also: *The Handbook*), edited by Ian James Kidd, José Medina and Gaile Pohlhaus, Jr. Various scholars, including the field's salient voice Miranda Fricker (and incidentally, Lorraine Code) updated, reviewed or topicalized their developed theories for this book. The featured articles analyze and clarify examples from historical to contemporary practices across a very broad range of contexts, which helped to train the necessary sensitivity to epistemic injustice dynamics across domains. In the book, Pohlhaus explains how the analysis of these injustices, "as a mix of political, ethical, and epistemological philosophical endeavors affords an understanding of how our institutional social arrangements can cramp living epistemic values like truth, aptness and understanding."²⁴⁸

246 Williams, *Truth and Truthfulness*.

247 Kidd, Medina, and Pohlhaus, Jr., 'Introduction to The Routledge Handbook on Epistemic Injustice'.

248 Gaile Pohlhaus, Jr., 'Varieties of Epistemic Injustice', in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017), 13; The field's examples in fact go well beyond cramp. In the worst varieties, epistemic injustice leads to hermeneutical death, and epistemicide. Medina, 'Varieties of Hermeneutical Injustice', 41, 47; Bar-Izhak et al, 'In Search of Epistemic Justice: A Tentative Cartography'.

3.1.2 Application of the consulted theories

The thesis's focus is on the duties of explainers and the rules that determine their duties. It was argued that explainers inevitably perform a function in the (further) production or prevention of 'harm by knowledge,' through their justifications of decisions. The larger research question (*"in light of the existence of wrongful knowledge practices; understanding the explanation of decisions to decision subjects as 'knowledge making about knowledge making,' and understanding rules that govern explanation as regulating 'conduct about conduct,' how can explanation rules promote responsible explainer behavior?"*) breaks down into four sub questions, the first two of which will be answered in this chapter.

Question 1, "how do knowledge practices produce and avoid to produce harms?" is answered by a descriptive analysis based on consulted literature. Section 3.2.1 introduces three key concepts, sections 3.2.2 to 3.2.4 parse insights that can usefully inform explanation norms into three categories: 'bad,' or how injustices are established and what they do to people; 'murky,' how bads get to be sustained in and by society; and 'good' about values to engage with to prevent and repair this. The categorization from bad to good is done to ensure the good is informed by the bad. The order helps to get acquainted with the dynamics, and understand how these influence, and are influenced by different 'knowledge' spheres and levels: local, societal, institutional. In reality, 'bad' and 'murky' will be hard to distinguish.

For question 2, "how can explainers avoid to re/produce epistemic harms their practices?" more insights are added, and the earlier insights made to count in the epistemic relationship this thesis focuses on: the institutionalized individual explanation relation. First, section 3.3.1 presents explanation as a type of interactive, testimonial practice that should 'demonstrably aim for' just treatment. It understands explanation as knowledge making about knowledge making, which involves conduct about conduct. I adds insights that see to the chosen setting and social interaction that takes place within it. Section 3.3.2 summarizes takeaways that are modeled into a set of four duties of care in section 3.3.3. These each describe behaviors for four phases of a comprehensive explanation cycle. In reality the phases will overlap and intertwine, but that is not problematic. The point is to raise awareness of the types of things that need to happen, and how each phase necessarily builds on, and eventually feeds back into, the others. The Model will be used to perform explanation domain research in following chapters, and can be used to inform explanation regulation more broadly. The tool is intended to be further developed through such uses.

3.2 Injustice and justice in knowledge practices

3.2.1 Key concepts: knowledge, justice, and truth

This section introduces (not: defines) the thesis's understanding of three key, recurring concepts whose colloquial and scientific understanding are too diverse (and sometimes contested) to 'simply use' when writing for an interdisciplinary audience. These are knowledge, understood as a product of sociality; justice, as respect for fundamental human inter-dependence and equality; and truth, understood as something that needs to be aimed for (but) with explicit justice-oriented aims. With this, this introductory section also introduces the consulted fields of research. AI developments are related to at times, but they are not the focus.

3.2.1.1 *Knowledge as a product of sociality*

There are types and items of knowledge about which it has been claimed that they are 'neutral.' With this it is meant that they do not express human normativity but only gain this dimension through how they are used. Typically, such arguments are made about the yield of the natural and quantitative sciences, but also about 'raw' data that algorithms are trained on.²⁴⁹ The neutrality argument has also been made about technology, including algorithms and AI, i.e. knowledge systems. And just like the argument fails for technology it fails for knowledge. Both are made by humans and involve making decisions, i.e. choices between theoretically unlimited options. Such choices (and choices about how to make choices) are inevitably normative, or they wouldn't matter.²⁵⁰ With regard to knowledge and its practices, philosophers have first had to argue how knowledge is social, after traditional epistemologies ignored, or refused to acknowledge this.²⁵¹ Lorraine Code for example writes how the sociality of knowledge expresses in how people necessarily interact with each other to seek and get to knowledge: they are interdependent and learn to be so from childhood onwards.²⁵² Knowledge is shaped through these interactions, and inevitably subject to the attitudes of participants. One step further, those concerned with epistemic justice study how power relations enacted through our sociality matter for the fairness of our epistemic lives: something that is not naturally recognized or acknowledged by all social epistemologists.²⁵³

249 Lisa Gitelman, ed., *Raw Data Is an Oxymoron* (MIT press, 2013).

250 Cohen, 'The Ethical Basis of Legal Criticism'.

251 Mitova, 'A New Argument for the Non-Instrumental Value of Truth', 9; Mills, 'White Ignorance', 230.

252 Code, *Epistemic Responsibility*, 169.

253 "[A]t least one major reason for this failure is that the conceptions of society in the literature too often presuppose a degree of consent and inclusion that does not exist outside the imagination of mainstream scholars—in a sense a societal population essentially generated by simple iteration of that originally solitary Cartesian cognizer (...) The concepts of domination, hegemony, ideology, mystification, exploitation, and so on that are part of the lingua franca of radicals find little or no place here. In particular, the analysis of the implications for social cognition and social ignorance of the legacy of white supremacy has barely been initiated." Mills, 'White Ignorance', 231.

That our sociality, normativity and power relations are of influence on our epistemic produce is well established and by now uncontested by many in the scientific community. Yet arguments that algorithmic technologies and (other) quantitative knowledges are inherently neutral tools are still made. Such denials of the normative dimensions of knowledge are highly problematic. Consider how ADS for the recognition and or prediction of human danger, harm or aggression (problematic in themselves in how many systems falsely claim to be able to gauge persons' state of mind from their facial expression)²⁵⁴, feed on concepts such as (the need to enable) 'self-defense' that are anything but neutral themselves already. Problematic histories where it is established who are the "grievable" selves that deserve to be protected against those who are defined as others deserve attention: whose understandings are promoted to a status deserving of protection and defense, and whose views get to be dismissed?²⁵⁵ In the US, the life-or-death importance that such knowledge claims have is made painfully clear each time a life is wrongfully lost and the unequal battle for the recognition of wrongfulness plays out in court.²⁵⁶

The example helps to see how where human morality and sociality are a factor, politics are involved: negotiation, discussion, conflict, power play, "complex strategic relations that shape lives."²⁵⁷ At the time of writing, a Dutch group of peaceful protesters against a yearly national 'Blackface' tradition²⁵⁸ were violently attacked by locals: petrol was poured over the car, doors were yanked open, possessions destroyed. The police stood by and did nothing short of alerting the Mayor of the nearby town to cancel the peaceful protest.²⁵⁹ When the only Black member of Parliament brought this racist response to the fore, the Minister of Justice responded fiercely by accusing her of wrongfully attacking the police force with racist accusations²⁶⁰—only months after the same Minister had acknowledged the existence of persistent racism in the Dutch police force in the same House of Parliament.²⁶¹

254 'Feldman Barret et al., Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements', *Psychological Science in the Public Interest* volume 20, issue 1, 2019.

255 Butler, *The Force of Nonviolence*, 9.

256 Well-known cases include the 1992 killing of Yoshihiro Hattori, 16 years old, the 2012 killing of Trayvon Martin, 17 years old, the 2020 killing of Ahmaud Arbery, 25 years old; the actual list is very long. Trayvon Martin's killing and the subsequent acquittal of his murderer inspired Alicia Garza, Patrisse Cullors, and Opal Tometi to start the Black Lives Matter movement. <https://blacklivesmatter.com/herstory/>.

257 Bacchi, 'Why Study Problematisations?', 1.

258 For an epistemic injustice take on this particular Dutch conflict, see T. J. Lagewaard, 'Epistemic Injustice and Deepened Disagreement', *Philosophical Studies* 178, nr. 5 (1 May 2021): 1571–92.

259 Amnesty International, 'Aangifte bedreiging en vernieling Sint-intocht Staphorst', 21 November 2022, <https://www.amnesty.nl/actueel/amnesty-nederland-doet-aangifte-van-bedeiging-rond-demonstratie-in-staphorst>.

260 The Minister also referred to the attackers as 'counter protesters,' a claim she later retracted. NOS.nl, 'Yesilgöz botst fel met Simons over nasleep Sint-intocht Staphorst', 22 November 2022, <https://nos.nl/artikel/2453437-yesilgoz-botst-fel-met-simons-over-nasleep-sint-intocht-staphorst>.

261 Ministerie van Justitie en Veiligheid, 'Antwoorden Kamervragen over discriminatie en racisme bij de politie', 6 July 2022; The discussion was fueled after a documentary about this was aired on Dutch television. *De blauwe familie*, 2022, <https://www.2doc.nl/documentaires/2022/05/de-blauwe-familie.html>.

The concept of self-defense's inherent reference to violent action against whoever is seen as an aggressor illustrates the importance of understanding the role of common knowledge spheres. People don't get through a single day without acting on what they accept as common knowledge, and so it is important to learn to understand what to have second thoughts about: what knowledge they can rely on safely for their own and others' sake. Ideally, 'their own' and 'others' conflate, and everyone shares their interests in qualifying what knowledge is safe to act on. In reality, acting on notions of what constitutes trespassing and may trigger acts of self-defense will be a totally different thing for different groups of people.

When making decisions about what knowledge to rely on, decisions about *who* to rely on are inevitably made as well, since knowledge is made by people. The perceived reliability of a source, then, is a factor in how people accept something as true(-ish). And just like it is people who set criteria for assessing knowledge and how it is made, people set criteria for the reliability of those who make knowledge claims: for their epistemic authority. Again, this brings in politics, and this will be discussed in much more detail in later sections.

Lastly, knowledge sits in a network of related terms such as belief, and truth. The thesis did not thoroughly engage with the vast bodies of work on all these terms' conceptualizations, although they certainly elicit useful scholarly discussion and disagreement. The exception is truth, which *is* dealt with to a limited extent later in this section. The reason for that is that the label 'knowledge' generally communicates that something is reliable, or true, for everyone, put on offer for being used and shared.²⁶²

3.2.1.2 *Justice: respecting and promoting fundamental equality of inter-dependent humans*

When seeking, creating, using, and sharing knowledge, those involved are driven by attitudes, and attitudes are value-laden. Justice-related values are among those that can be engaged with while 'doing knowledge.' The justice that this chapter zooms is *about* knowledge. This can be perfectly aligned with other notions of justice (e.g., distributive, social, environmental), or it can inform them. The justice that is investigated and applied to institutionalized explaining starts from the understanding of humanity as a global knowledge community whose members are fundamentally equal, and fundamentally inter-dependent. This means that knowledge practices should not exclude anyone (or any group) on wrongful grounds, or wrongfully serve some groups better than others. The need to strive for inclusivity and diversity in knowledge practices and their produce is a big topic of critique on the AI knowledge making community.

262 Amore alerts to the fact that machine learning knowledge claims in fact signal a different kind of 'true,' a 'this might happen,' accompanied by a probability percentage for users of a system to inform their decisions. This dynamic feeds arguments that ML system's potential for harm lies solely in their use. Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*.

Theoretically, acting in service of only a part of humanity is impossible because eventually, all will be harmed by the cultivation of bad knowledge; environmental knowledge is a case in point. But ‘bad knowledge’ can harm certain groups for long periods of time while, and because, other groups benefit from it. And as introduced in the previous section, people who benefit may become unaware of the fact that what they accept as common knowledge is unfair to other groups of people. The example illustrates how acting in service of justice-for-all when ‘doing knowledge’ entails cultivating various attitudes and dispositions of awareness (historical, scientific, reflective ..), and a preparedness to promote proper practices and object to bad ones. It also sustains the choice to make sure that idealizations (‘the good’) are informed by the non-ideal.

In Code’s terms, principally acting towards the benefit of all in knowledge practices is called ‘responsibility.’ Finding out what it means to be an epistemically responsible actor entails to understand the “manner of [knowledge’s] mattering and the implications thereof.”²⁶³ For this thesis, justice and her description of responsibility overlap, and they will be used interchangeably. Williams argues to understand the difference between striving for freedom *from* humanity in knowledge practices, and striving for freedom *for* humanity. Although there is merit in trying to ‘liberate’ a knowledge seeking endeavor from particular human interests and intent, this simply means that a choice is made for certain interests over others. The point therefore should be to seek maximum freedom *for* the whole of humanity. This is done by avoiding, and destroying, socially oppressive representations in knowledge produce.²⁶⁴ Practices, in other words, need to avoid to be wrongful so as not to produce harms.

3.2.1.3 *Truth: the merits of chasing non-oppressive representations*

There is a type of ‘relativism’ that misuses arguments about the diversity of knowledge and its practitioners to stop striving for precisely what Williams argued for: knowledge in service of all people. In such cases, knowledge gets to be defended as ‘true for us,’ immune from attacks from others who would like to see their interests represented; or dismissed as ‘true for them,’ but not up to ‘our’ standards of knowledge. Such relativism is the opposite of the understanding of truth that this thesis adopts.

Williams and Mitova both argue for truth as having a non-instrumental, yet ‘extrinsic’ value in how it is the *aim* of inquiry, and inquiry is valuable in itself. Mitova describes inquiry as a cognitive activity aimed at “finding out how things stand on a particular topic,” an activity that should be available to all epistemic agents.²⁶⁵ She stresses how this understanding is different from describing truth as the *product* of inquiry,²⁶⁶ and

²⁶³ Code, *Epistemic Responsibility*, 132.

²⁶⁴ Williams, *Truth and Truthfulness*, 184.

²⁶⁵ Mitova, ‘A New Argument for the Non-Instrumental Value of Truth’, 17.

²⁶⁶ Mitova, 23.

herein lies the key to keeping ‘truth’ open-ended in order to be able to contribute to the improvement of truth-claims. Truth as such is a statement about something in the world, a claim that is meant to be ‘solid enough’ to symbolize something ‘real enough’ to rely on, but that still needs to remain fundamentally contestable because nothing ‘out there,’ object or subject, imposes its meaning. Whatever it is that people find out about the world is co-determined by how they seek it.²⁶⁷ The truth of this for the scientific practice of truth finding is obvious; it would be useless to teach and publish scientific findings only.²⁶⁸ Williams warns academics to “take care, and do not lie,”²⁶⁹ where ‘take care’ refers to investigative (methodological) choices, and ‘do not lie’ to the honesty in presentations of truth claims.

In service of the kind of justice aimed for, there is a need to fight oppressive definitions of what is true (e.g., certain kinds of people are less intelligent) without reducing the status of *all* knowledge to ‘entirely relative’; to *mere* social agreement;²⁷⁰ to ‘beliefs’ when claimed by those who don’t make the rules about what counts.²⁷¹ Such relativism impedes the possibility to argue against oppressive meanings on the basis that something is *different in reality* than it is presented to be. Progress and innovation, including social progress and innovation (if there is a difference) would stifle without the ability to do so. In other words, there is value in *striving* for truth because of what the striving makes us do,²⁷² but also in keeping truth ‘out there’ as something to strive *for*: a reality that we need to seek agreement about. In that sense, truth resembles objectivity: it does not exist, but as Mills argues, it needs to be the ideal to pursue in order to fight the wrong kinds of (relativist) subjectivity.²⁷³

267 Marx W Wartofsky, ‘What Can the Epistemologists Learn from the Endocrinologists? Or Is the Philosophy of Medicine Based on a Mistake?’ in Ronald A Carson and Chester R Burns (eds), *Philosophy of Medicine and Bioethics: A Twenty-Year Retrospective and Critical Appraisal* (Springer Netherlands 1997) 64.

268 Edmund D. Pellegrino, ‘Praxis as a Keystone for the Philosophy and Professional Ethics of Medicine: The Need for an Arch-Support: Commentary on Toulmin and Wartofsky’, in *Philosophy of Medicine and Bioethics: A Twenty-Year Retrospective and Critical Appraisal*, edited by Ronald A. Carson and Chester R. Burns, Philosophy and Medicine (Dordrecht: Springer Netherlands, 1997), 78.

269 This characterization specifically was ‘for academics,’ Williams, *Truth and Truthfulness*, 11.

270 Geuskens Epistemic Justice: A Principled Approach to Knowledge Generation and Distribution’ (Tilburg University, 2018).

271 Lawrence Bonjour, “Can empirical knowledge have a foundation?”, *American Philosophical Quarterly* 15:1 (1978), in: Lorraine Code, ‘Epistemic Responsibility (2017)’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017).

272 Mitova, ‘A New Argument for the Non-Instrumental Value of Truth’.

273 Mills, “‘Ideal Theory’ as Ideology”, 174.

When inquiry is valuable in itself, and knowledge practices can do justice to persons or the opposite, the decision to accept a claim as solid enough, allowed to represent something real enough, should also depend on how the claim was negotiated.²⁷⁴ To cite Grasswick on injustices in, and through, scientific knowledge making, “[i]njustices deal in social relations and interactions. Epistemic injustices exist because a large portion of our epistemic lives are social.”²⁷⁵ The thesis investigates how such notions are, or should be, represented in how we regulated explanation: as explanation and justification of decisional processes, and as knowledge practices themselves in which the truth of a decisional claim is negotiated. Think of legal rules as ‘ground truths’ (which algorithms are now making up themselves), of how information about people is used as truth that decisions are based on, and consider that explanations are referred to as truthful accounts of decisional processes. An objection may be brought up here with regard to legal truths in particular. Most truth pretense in legal decisions is not so much a solid claim about the world (let alone an ‘open ended’ one) but rather an agreement about facts, made according to agreed upon procedures. Negotiated conclusions that allow parties to move on from situations that would remain unresolved otherwise (or not in ways that we find acceptable). But these practices too can certainly be investigated in terms of how ‘just’ such negotiations are.²⁷⁶

A brief return to the example of self-defense helps to see how science, law, and politics; knowledge, justice and truth interact. The legal right to self-defense sustains truth-claims about the justness of killing a person. It therefore remains (ever) necessary to investigate both the *concept* of self-defense in law for how it is wrongly skewed towards the safety of some at the cost of others, as well as the truth claims

274 Foucault, cited by Allen: “[m]y problem is to know how men govern (themselves and others) by means of the production of truth.” Amy Allen, ‘Power/Knowledge/Resistance: Foucault and Epistemic Injustice’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017).

275 Grasswick, ‘Epistemic Injustice in Science’.

276 Sullivan argues how “the general trajectory” of epistemic injustice’s diverse and interdisciplinary discourse “can be read as a call to raise critical consciousness regarding the shortcomings of our truth-finding and legal practices.” Michael Sullivan, ‘Epistemic Justice and the Law’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017), 295; An interesting example are truth-and-reconciliation commissions. These fact-finding bodies are established in times of transitional justice after grave human rights infringements, such as after civil war, or the well-known example at the official end of Apartheid in South-Africa. Through various methods, a legally stamped truth about what happened is negotiated so that people can move on with at least an acknowledgment of what happened to them and in the same place with the perpetrators of it, even if there is no possibility (yet) to qualify what these truths should mean in terms of accountability and consequences. Such commissions are also criticized, for example in how they short-stop the kind of truth finding that a community needs in order to heal and for imposing forgiveness. Legal proceedings after grave ‘civil’ infringements may fail just as well. In both cases truths are buried but continue to foster, leading to mistrust in between members of communities among other things. Natasha Stamenkovikj, ‘The Truth in Times of Transitional Justice: The Council of Europe and the Former Yugoslavia’ (Tilburg, Tilburg University, 2019); Monica Black, *A Demon-Haunted Land: Witches, Wonder Doctors, and the Ghosts of the Past in Post-WWII Germany* (New York, New York: Metropolitan Books, 2020).

that are built on these understandings (the justness of killing a person). Contemporary AI methods (science) are making the honest representations (“take care, and do not lie”) of both kinds of truth-claims increasingly hard. ML systems obscurely categorize persons’ behaviors and movements as ‘threatening,’ and attach a risk label. These are “*ethicopolitical*” claims,²⁷⁷ that come to influence how people see and deal with the world. As we saw, how this compromises explanation is accepted, less clear is to what depth of understanding of their ‘knowledge systems’ explainers are expected to engage with more generally.

In the next three sections, the above-described understandings of knowledge, justice, and truth are used in discussions of how injustices are established and what they do to people (bad), how they get to be sustained by society (murky), and what values to engage with to prevent and repair this (good).

3.2.2 Bad: direct harms as an effect of misconstruction and misuse of epistemic authority

Decision makers and explainers hold positions of social (decisional) and informational authority in relation to their explainees. These two types of authority relate and combine as ‘social epistemic authority.’ What this means is important to understand, and a thought about primary relations (children and who raises them, teachers and pupils) helps to start thinking about this. In these primary relations, the intertwining of social and informational authority is a given. Homes and schools are where people first learn to rely on knowledge and the people who tell and teach it. Put differently, where they learn to be *inter*-dependent, rather than just *in*-dependent thinkers and actors.²⁷⁸ These primary relations, for better or for worse, prepare children for relations later on in life, relations where they again need to relate to others who are authoritatively ‘in the know,’²⁷⁹ and how and when to trust their interactions with them. Think of higher education, the need to consult a doctor, or a civil servant for support.

Ideally, the persons that they encounter are put in positions of social-epistemic authority for good reason. In reality people also end up in such positions for reasons unrelated to the quality of their expertise. Ideally, the people they encounter are able and inclined to treat them with respect as cognitive agents, but this is not a given either and they may experience injustices of various kinds. Terms much used to describe types of wrongs and harms that people experience are testimonial and hermeneutical injustice. Many authors build on Miranda Fricker’s famous conception of these notions²⁸⁰ and Fricker’s name has become closely associated with epistemic injustice

277 Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, 17.

278 As, rather than (just) independent, it’s where we learn to be interdependent, Code, *Epistemic Responsibility*.

279 Code, 60.

280 Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, *Epistemic Injustice* (Oxford University Press, 2007).

itself as a concept and a term. Testimonial injustice describes harms inflicted people as knowers, where their knowledge, and knowledge related capabilities are held in low regard for unjust reasons. People are for example treated as less intelligent, otherwise incapable, or less trustworthy. This can be flagrant and explicit or under the radar, when it for example expresses as implicit bias: implicit cognition of influence “on our behaviors and judgements, contributing to patterns of discriminatory behavior.”²⁸¹ The strong connotation of the term ‘testimony’ with courtroom procedures is not entirely unhelpful: think of reduced credibility of witnesses and defendants from marginalized groups in criminal procedures.²⁸² But it should not limit the understanding of other ways and places that testimonial injustices occur, or when the term itself (in its current understanding) lacks conceptual depth. E.g., people who suffer from reduced credibility or recognition of their capabilities are likely to experience reduced opportunities in terms of education and career.²⁸³ This reduced ability to ‘achieve’ feeds into the same wrongful perceptions that put them in this spot in the first place.²⁸⁴ For an example of a type of epistemic injustice that *could* be called testimonial but is better served with a term of its own: Mitova introduces the concept ‘explanatory injustice’ for situations where acknowledgement of *agential* authority (expressing in ‘motivating reason’) is wrongfully withheld. This happens when x-ist (and frequently immoral) ‘reasons why’ are ascribed to the actions and utterings of members of the wronged groups.²⁸⁵

It was already discussed how the use of correlative methods for the assessment of people in all such situations (grading, diagnosis, eligibility, jobs, courtrooms) leads to the exacerbation of such wrongs. The point (of course) is not to blame technology, but to illustrate the importance of understanding the justness, and justness potential, of the knowledge practices that inform and sustain decisional practices. Code voiced concerns about scientific practice specifically, as a practice that does not just define its own goods and methods, but also controls the ‘channels’ by which contributions are submitted, making it resistant to the diversification of participants. When these are populated with persons who are mainly unaffected by, uninterested in, or unacquainted with the unjust consequences of the knowledge they produce, they become places where bad ‘goods’ persist, and are internalized by its participants.²⁸⁶ Benjamin warns

281 For an interesting argument about responsibility with regard to maintaining biases, see Jules Holroyd, ‘Responsibility for Implicit Bias’, *Journal of Social Philosophy* 43, nr. 3 (2012): 274–306.

282 E.g., the existence (in jury trials) of juror’s biased perceptions of testifiers propels Sullivan to argue jurors should improve their “critical self-awareness,” for example via implicit bias tests: Sullivan, ‘Epistemic Justice and the Law’ For such a test, see <https://implicit.harvard.edu/implicit/takeatest.html>.

283 Kotzee, ‘Education and Epistemic Injustice’.

284 ‘But as feminist, race and queer-theorists have taught us, power is often a zero-sum game. If someone’s empowered in the social domain, it is typically at the expense of someone else.’ Veli Mitova, ‘Explanatory Injustice and Epistemic Agency’, *Ethical Theory and Moral Practice* 23, nr. 5 (November 2020): 8.

285 ‘[C]ertain reason-ascriptions empower while others disempower, and (consequently) how through them believers can be belittled as epistemic agents.’ Mitova, 2,9.

286 Code, *Epistemic Responsibility*, 231.

how knowledge making authority needs to be *fundamentally* diverse and inclusive enough to avoid patterns of dominance to persist, perchance hidden by practices dressed up to either be diverse, or serve diversity's aims.²⁸⁷

The example serves as a bridge to the second type of injustice named above: hermeneutical injustice, which is a more complex concept. This type of injustice describes harms done to people's *capabilities* for meaning making, and meaning sharing: that which needs to happen before a claim can even be made, regardless of whether it is then dismissed through testimonial injustice. An example from legal knowledge making: upon retirement, the UK's Justice Lady Hale was asked to name three 'desert island judgments'.²⁸⁸ One of her picks was a 2011 judgment that broadened the legal definition of domestic violence to include violence of the 'non-physical' kind. Hale: "We said, no, [domestic violence] involves all kinds of domination and abuse. Psychological domination – what we now call coercive control. We didn't have that phrase then."²⁸⁹ So until then, what was arguably a wrongful *exclusion* of this type of violence from the definition that was explained in judicial rulings did not just wrong those who experienced it, but also harmed them by providing no term to trigger protection with. Such updates were named earlier as examples of 'progress,' but consider that this particular one took until 2011 to establish. Consider also that whether progress *can* establish via judicial routes and in a reliable way, is dependent on political choices about the larger institutional system. At the time of writing, the US Supreme Court struck down the seminal 1973 *Roe v Wade* ruling that gave women the right to abortion,²⁹⁰ which itself is part of a set of rulings that together establish a range of reproductive and family life rights (notably LHBTQI rights, contraception).²⁹¹ The dissenting Justices argued how the ruling is a direct political-ideological interference, not based on actual 'progress' in societal discourse on the subject. If anything, abortion has become *more* fiercely debated from both sides, which would call for judicial restraint and a maintenance of the status quo in which a right has been established 50 years ago.²⁹² The ruling is welcomed by politicians thus ideologically inclined, and

287 Benjamin, *Race After Technology*, 144–49 In the case of AI, even the practice's critical scholarly circles are criticized as being non-diverse. See e.g., the statement on 'diversity and inclusion' statement for the 2020 Fairness, Accountability and Transparency in Computing conference 2020: <https://facctconference.org/2020/inclusion.html>.

288 *Yemshaw v London Borough of Hounslow*, on appeal from 2009 EWCA Civ 1543, UK Supreme Court, 2011 Simon Hattenstone, 'Lady Hale: "My Desert Island Judgments? Number One Would Probably Be the Prorogation Case"', *The Guardian*, 11 January 2020, sec. Law, <https://www.theguardian.com/law/2020/jan/11/lady-hale-desert-island-judgments-prorogation-case-simon-hattenstone>.

289 Hattenstone.

290 19-1392 *Dobbs v. Jackson Women's Health Organization* (06/24/2022) (US Supreme Court 2022).

291 Jessica Glenza, 'How Dismantling *Roe v Wade* Could Imperil Other "Core, Basic Human Rights"', *The Guardian*, 11 December 2021, sec. US news, <https://www.theguardian.com/us-news/2021/dec/11/supreme-court-ro-v-wade-gay-rights-contraceptives-fertility-treatments>.

292 "If the Court should make a choice at such times, it is established that a majority of US citizens would like to see 'Roe' upheld." "Fewer Rights than Their Grandmothers": Read Three Justices' Searing Abortion Dissent', *The Guardian*, 24 June 2022, sec. Opinion, <https://www.theguardian.com/commentisfree/2022/jun/24/supreme-court-ro-v-wade-breyer-sotomayor-kegan>.

they are eager to interfere with women's cognitive agency. The Governor of South Dakota expressed this clearly: "Every abortion always had two victims: the unborn child and the mother. (...) We must do what we can to help mothers in crisis know that there are options and resources available for them. Together, we will ensure that abortion is not only illegal in South Dakota – it is unthinkable."²⁹³

The ideological interference puts US medical decision makers in a jam. The principle of 'shared decision-making' has become the norm, and requires doctors to engage their epistemic authority only in service of the personal values of their patients. In other words, "[t]here is no room within the sanctuary of the patient-physician relationship for individual lawmakers who wish to impose their personal religious or ideological views on others."²⁹⁴ But this presentation of the ideal principle denies the rather ugly truth that medicine was always a salient locus of epistemic injustices. Examples from this domain illustrate how the two types of injustice are intimately related to the point that disentangling them may be neither possible or useful.²⁹⁵ Women have for example been widely excluded from, and ignored in medical research and the development of diagnostic methods, harming their health.²⁹⁶ Depriving them of disease or treatment *vocabulary* is nothing new, also for other discriminated groups.²⁹⁷ The lack of a means to express complaints leads to further maltreatment, especially for those who already suffer from reduced credibility, or testimonial injustice – intersectional dynamics apply,²⁹⁸ and the categories conflate. Chapter 5 analyzes Dutch GP explanation rules for how they bear on medical knowledge practices.

When such situations and practices of epistemic injustice are allowed to continue undisturbed, the growing historical distance to origins and sources of a situation makes it harder to get retrospective insights, and to make them land and take root in common contemporary awareness. This is the subject of the next section. The point for here is how this harms people quite directly. No-one is immune to becoming familiarized with the world they grow up in, including any accepted truths in it that represent them unfairly. E.g., gendered attributions of emotional instability and physical weakness, negative relations between intelligence, benevolence, and ethnicity lower the self-

293 'Gov. Noem and Legislative Leaders Announce Plans for Special Session to Save Lives, Help Mothers', South Dakota State news website, consulted on 25 June 2022, <https://news.sd.gov/newsitem.aspx?id=30323>.

294 Cited is Dr Ifthah A Hoskins, president of the American College of Obstetricians and Gynecologists Jessica Glenza and Martin Pengelly, 'US Supreme Court Overturns Abortion Rights, Upending Roe v Wade', *The Guardian*, 24 June 2022, sec. World news, <https://www.theguardian.com/world/2022/jun/24/roe-v-wade-overturned-abortion-summary-supreme-court>.

295 As classifications do, they help to clarify, but they should –and do– also provoke critique and further development. Medina, 'Varieties of Hermeneutical Injustice'.

296 Maya Dusenbery, *Doing Harm: The Truth About How Bad Medicine and Lazy Science Leave Women Dismissed, Misdiagnosed, and Sick* (HarperCollins, 2018).

297 Carel and Kidd, 'Epistemic Injustice in Medicine and Health Care'.

298 Kimberle Crenshaw, 'Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color', *Stanford Law Review* 43, nr. 6 (1991): 1241–99.

esteem of those affected, and how they regard their peers and those in favorable positions. Congdon builds on recognition theory, which explains how we need others to maintain positive relations to ourselves.²⁹⁹ Persistent attributions of negative traits and behavior in absence of proper recognition (consciously and unconsciously) also influences actual behavior, producing additional ‘evidence’ of what *still* is untrue. Liebow calls attention to how where this ‘allows’ some persons to conform and fit in, for others there is no escape: “the harm of internalizing stereotypes about criminality for people of color .. positions agents to view themselves as existing outside of the moral community and provides few entry points back into the moral fold.”³⁰⁰

3.2.3 Murky: perpetuation of wrongs in and through shared knowledge spheres

3.2.3.1 The importance and influence of science as a knowledge sphere (revisited)

This section zooms in on social-epistemic dynamics on community and societal levels. The point is to understand more about how injustices persist in communal knowledge spheres that determine the information positions of explainers and their explainees. This time the discussion starts with a modern, AI-related example. Medical scientists recently voiced critique on the lack of insight that tech companies provide in their publications in medical scientific journals. They warn how “[m]any egregious failures of science were due to lack of public access to code and data used in the discovery process.” Publicizing ‘mere’ results, as they argued such behavior amounts to, comes down to promoting “closed technology.”³⁰¹ The critique was widely publicized, also outside of the medical fields.³⁰² The original (tech) article had still passed peer review, which is concerning. The question is whether the condemnation testifies to the scientific system’s ability for progressive uptake.

It is tempting to think of bad quality knowledge as self-destructive. If not via self-regulation, like in the example above, then on the basis of the more “cynical”³⁰³ argument that eventually, the produce of low quality, non-inclusive knowledge paradigms will harm everyone. At that point the groups that used to benefit lose their reasons to persist their epistemic oppression. But aside from the unacceptable costs of waiting, there is no evidence that this will happen at all. The ideal ignores the fierce resistance to critique and change within knowledge communities, *especially* those who

299 Matthew Congdon, ‘What’s Wrong with Epistemic Injustice? Harm, Vice, Objectification, Misrecognition’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017), 248.

300 Nabina Liebow, “Internalized Oppression and Its Varied Moral Harms: Self-Perceptions of Reduced Agency and Criminality,” *Hypatia* 31, no. 4 (2016): 723.

301 Benjamin Haibe-Kains et al, ‘Transparency and Reproducibility in Artificial Intelligence’, *Nature* 586, nr. 7829 (October 2020): E14–16.

302 ‘AI Is Wrestling with a Replication Crisis’.

303 Benjamin, *Race After Technology*, 76.

“define [their] own goods and methods,”³⁰⁴ Notions of guilt, responsibility,³⁰⁵ and the justice of providing reparations for past wrongs tend to be resisted by those who don’t see how a system privileges them. Kuhn famously argued how scientific paradigms (are paradigms, and) are very strong belief systems.³⁰⁶ This is particularly problematic for a practice whose core business it is to create reliable knowledge in responsible ways. Science’s social-epistemic authority is generously awarded in societies who value “scientific ways of knowing,” such as ours.³⁰⁷

Critical scholars revealed the shaping influence of many oppressive ideologies in, and on, scientific practice.³⁰⁸ Examples abound of scientific practices that are co-opted by, and or acting as, a force of oppression. Effects include exclusionary knowledge cultures, ethical and physical abuse of human research subjects, misguided research programs, and self-proving, behavioral science-oriented practices.³⁰⁹ The wrongful knowledge produced this way both derives its authority from, and grants authority to, unjust notions at play in populations where the findings land. The history of ‘genetics’ is a case in point.³¹⁰ In 1994, *The Bell Curve* linked ‘race,’ via genetics, to intelligence. The book’s authors excluded salient developments and arguments from all the scientific bases that grounded their arguments. By the time it was published, broad scientific agreement had already been reached about the lack of hereditary relations between the categories, as well as on the impotence of the categories themselves. Established (causal) relations between financial resources and academic success, including success rates for IQ testing, further undermined the writers’ argument.³¹¹ The general book-buying population did not care: the book became a best-seller.

304 Code, *Epistemic Responsibility*, 231.

305 Arendt’s work on notions of guilt and responsibility also helps to understand this. Hannah Arendt, ‘Some Questions of Moral Philosophy’, *Responsibility and Judgment* (Reprint edition, Schocken 2005).

306 Thomas S Kuhn, *The Structure of Scientific Revolutions* (University of Chicago Press 1962).

307 Grasswick, ‘Epistemic Injustice in Science’, 313, 321; see also Garland E. Allen, ‘The Ideology of Elimination: American and German Eugenics, 1900-1945’, in *Medicine and Medical Ethics in Nazi Germany: Origins, Practices, Legacies*, edited by van Francis R. Nicosia and Jonathan Huener, 1st Edition (New York: Berghahn Books, 2002), 14 This not to ignore the contemporary attacks (virtual and actual) on Scientists and ‘scientific ways of knowing.’ Later on in the thesis, this problem and how to understand its complex origins is discussed in relation to the 2020 Coronavirus pandemic.

308 Grasswick, “Epistemic Injustice in Science,” 313.

309 An example that shows how analysis needs to include financial, narrative, and other powers to understand how this happens: Annelies Kleinherenbrink, ‘The Politics of Plasticity: Sex and Gender in the 21st Century Brain’ (Thesis, University of Amsterdam, 2016).

310 Allen, ‘The Ideology of Elimination: American and German Eugenics, 1900-1945’, 30.

311 In a critical race theory-informed ‘radical assessment’ of the questionable motives that underly the book’s publication, Bell imagines the authors in fact found white intelligence to be inferior. Wanting to protect black people from the lethal backlash these findings would produce, they published the acquiescing opposite. The account aptly and ironically analyzes some extremities of racist epistemic injustice. Derrick A. Bell, ‘Who’s Afraid of Critical Race Theory?’, *University of Illinois law Review* 1995, nr. 893 (1995).

It is fair to ask whether the book buyers awarded social authority to scientific ways of knowing, or to cherry-picked conclusions that justified their established social dominance. Thinking about this helps to understand that even when science does ‘self-correct,’ the corrections are not necessarily accepted, let alone embraced. Scientists’ “desire to discover and hold on to reality [which] can stand against such forces as political corruption and terror,”³¹² a disposition that Williams apparently awards with what seems like a lot of faith, is also not enough. Beckwith and Pierce analyzed how the uptake of untrue claims (‘criminal genes’, and other genetic determinants of bad behavior) in (pop)culture and even policy becomes highly resistant to negations of the claims even when they are issued by the original scientists.³¹³ The authors argue the merit of including ontologies of such wrongful dynamics in the curricula of trainee medical knowledge makers. The idea aligns with the broader argument that scientists should learn to “consider the political fallout of [their] theoretical investments.”³¹⁴ Green et al. argue that ethical considerations need to start when investments are still theoretical. There is necessary thinking to be done about which investigations and design aims *not* to pursue, but this is problematically off-table.³¹⁵

3.2.3.2 *Low maintenance oppression in common knowledge spheres*

When histories of racist, sexist, ableist and other wrongful ideology-driven knowledge practices fail to register as common knowledge, their societal uptake continues to foster. Sustained by neglect, cultural-political selectivity (see above), and other more and less deliberate forms of “ignorance making,”³¹⁶ low maintenance epistemic oppression is born. This point picks up from the previous section’s discussion about the influence of passing time. Wrongful dominance in a knowledge community becomes harder to identify when the roots of it lie generations away. It is a deliberate tactic of colonial powers to shape a knowledge society in a way that helps along such ‘forgetting.’ E.g., local knowledges are strategically dismissed, local knowers are excluded from epistemic participation in education, society, and politics.³¹⁷

312 Williams, *Truth and Truthfulness*, 143.

313 J. Beckwith and R. Pierce, ‘Genes and Human Behavior: Ethical Implications,’ in *Molecular-Genetic and Statistical Techniques for Behavioral and Neural Research*, edited by RT Gerlai (Elsevier Academic Press, 2018).

314 Bacchi, ‘Why Study Problematizations?’, 7; To be sure, this is not an easy task, and comes with its own ethical challenges Behnam Taebi, Jeroen van den Hoven, and Stephanie J. Bird, ‘The Importance of Ethics in Modern Universities of Technology’, *Science and Engineering Ethics* 25, nr. 6 (1 December 2019): 1625–32.

315 Daniel Greene, Anna Lauren Hoffmann, and Luke Stark, ‘Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning’, 2019, <https://doi.org/10.24251/HICSS.2019.258>.

316 Coined by Proctor as ‘agnotology,’ or “how ignorance is produced or maintained in diverse settings, through mechanisms such as deliberate or inadvertent neglect, secrecy and suppression, document destruction, unquestioned tradition, and myriad forms of inherent (or avoidable) culturopolitical selectivity”: Robert N. Proctor en Londa Schiebinger, ed., *Agnotology: The Making and Unmaking of Ignorance* (Stanford University Press, 2008).

317 Alcock, ‘Philosophy and Philosophical Practice: Eurocentrism as an Epistemology of Ignorance’.

The term ‘low maintenance’ is chosen to emphasize that dismantling, let alone deconstructing, unjust knowledge paradigms takes much time, many people’s persistence, and multi-disciplinary research efforts.³¹⁸ The previous section’s discussion of hermeneutical injustice helps to understand how a lack of vocabulary for wrongs is an obstacle for doing such work. The section also stated that the concept of hermeneutical injustice is complex. Medina argues that in order to identify and study epistemically unjust knowledge paradigms, investigations need to analyze expressive and interpretative resources, but also their sociology: how they are used, by whom, in what ways, with what effects.³¹⁹

For whoever is harmed by such deep-rooted epistemic injustices, fighting for change is especially exhaustive. It requires working under “excessive burdens of proof.”³²⁰ A seminal example is the concept of white ignorance,³²¹ described by Mills as the denial that racist ideologies got to shape whole social-epistemic systems. What in the past were explicit, and therewith *undeniable* purposes of domination have faded from awareness, sustained by an absence of felt urgency of privileged groups to know and understand. Dominant conceptual resources are broadly internalized, also by those at the receiving end of injustices. This exacerbates the harms of the unjust situation: the privileged group’s ignorance broadens to ‘shared reality bias’.³²²

In light of the unfair burdens of resisting, educating, and correcting, this section cites a bit of testimony from a privileged knower who referred himself to (pre-Musk) twitter with regrets about having “contributed to a coarsening society.”³²³ The white, male, US musician wonders how actually progressive their body of so-called progressive 1980s music was:³²⁴ music that audibly expressed the kinds of ‘traditional’ violence that it investigated. “If anything,” he explained in an interview about the tweets, “we were trying to underscore the banality, the everyday nonchalance toward our common history with the atrocious, all the while laboring under the tacit *mistaken* notion that things were getting better.”³²⁵ The insight that everyday nonchalance is a privilege

318 Pohlhaus, Jr., ‘Varieties of Epistemic Injustice’.

319 He also argues the category of hermeneutical injustices will need to be further developed, and designed a ‘starter kit’ of classifications to move along this work. The starter kit includes the following categories: source, dynamics, breadth, and depth of a problem Medina, ‘Varieties of Hermeneutical Injustice’, 43,45.

320 Margaret Urban Walker, ‘Truth Telling as Reparations’, *Metaphilosophy* 41, nr. 4 (2010): 525–45.

321 Mills, ‘White Ignorance’.

322 Anderson, ‘Epistemic Justice as a Virtue of Social Institutions’, 170.

323 Steve Albini (@electricalWSOP) on twitter October 12, 2021

324 The question corresponds with the questions about represented progressiveness that run as a red thread throughout the thesis. The author was well acquainted with the music, which at the time was considered to be on the cutting edge of progressiveness and exploitation.

325 “‘I’m Overdue for a Discussion About My Role in Inspiring ‘Edgelord’ Shit’: A Conversation with Steve Albini”, *MEL Magazine* (blog), 8 November 2021, <https://melmagazine.com/en-us/story/steve-albini-counsel-culture-interview>.

itself came later. “That’s the way a lot of straight white guys think of the world (..) The notion is that if you’re not actively doing something to oppress somebody, then you’re not part of the problem. As opposed to quietly enjoying all of the privilege that’s been bestowed on you by generations of this dominance.”³²⁶ The example bridges to the discussion in the next section, that engages with possible strategies toward prevention and restoration.

3.2.4 Good: societal and institutional promotion of preventative and corrective labor

3.2.4.1 Caring about trustworthiness: where methods and disposition meet.

Among scholarly concerns about how AI obfuscates important aspects of knowledge making is the need to maintain a “critical infrastructure.. a commons.. conducive to moral development and moral agency.”³²⁷ The previous sections warn not to prioritize any particular (scientific or other) communities’ choices with regard to what such critical infrastructures should look like. This section engages with notions about responsible, justice-oriented knowledge practices in the consulted literature. The discussions in this section increasingly ‘fit’ with the situation of institutionalized explanation that is the focus of the thesis.

To start with, different authors describe a tandem of values, dispositions, or virtues that that see to the making, sharing, and justifying of knowledge claims. They argue that these need to be cultivated to avoid injustice. They also see to the need to responsibly *understand*: to assess what is presented as knowledge and who is presenting it, which helps to think about the role authority of explainers themselves as explainees—which they inevitably also are. This chapter did not, and will not, precisely define the concept ‘virtue.’ Williams describes the general purposes of ‘virtue’ as “co-operation, self-transcendence, and social dignity.”³²⁸ Virtues in his argument are engaged in acts of truth-finding. Code defines virtues as principles, vehicles that need to remain open to discussion. She wishes to progress philosophical understandings of epistemic virtue which have not sufficiently centered ‘responsibility.’³²⁹ In what follows, the words virtue, trait, values and dispositions will be used interchangeably.³³⁰

326 “‘I’m Overdue for a Discussion About My Role in Inspiring ‘Edgelord’ Shit’”.

327 Roger Brownsword, ‘Law Disrupted, Law Re-Imagined, Law Re-Invented’, *Technology and Regulation*, 20 May 2019, 10–30; See also, Bygrave, ‘Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions’.

328 Williams, *Truth and Truthfulness*, 191.

329 Code, ‘Epistemic Responsibility (2017)’.

330 Readers who want to further explore the concept ‘virtues’ are referred to Heather Battaly, ‘Testimonial Injustice, Epistemic Vice, and Vice Epistemology’, in *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., paperback edition (Routledge, 2017); James A. Marcum, ‘Clinical Decision-Making, Gender Bias, Virtue Epistemology, and Quality Healthcare’, *Topoi* 36, nr. 3 (1 September 2017): 501–8.

Code's set of virtues are due care and intellectual honesty.³³¹ Williams (who does not cite her) uses accuracy and sincerity: "you do your best to acquire true beliefs, and what you say reflects what you believe."³³² Fricker writes about competence and trustworthiness.³³³ All conceptual pairs promote two, related motivations. The one leg of the tandem sees to the *creation* of knowledge *on* responsible (or 'virtuous') terms. The other to the motivation to *share* knowledge *in* responsible terms. Descriptions of the first leg primarily reflect back on methods, those of the second on the dispositions of practitioners. Importantly, the descriptions deliberately draw 'method' (or 'craft') and disposition, (or 'attitude') together.

Arguing for the first leg of the tandem (accuracy, due care, competence) Code is critical of how science promoted reliability over responsibility, and reliabilist rather than responsibilist dispositions.³³⁴ Her concern is with how reliabilism promotes universal, definite agreement on knowledge 'about the world' as attainable while failing to acknowledge that human attitudes drive any search for truth and 'warrant-ability.' Responsibilism, conversely, thrives on the premise that there is a quality to any one's understanding, a variable of the way the quest was undertaken.³³⁵ The description fits into Williams's argument to center the values of truth-finding over truth-claiming. I.e. any *pretense* of objectivity needs to be avoided and the aim of the (continued) search for non-oppressive representations needs to be explicit.³³⁶ These arguments are also arguments for the responsible estimation of truth claims. Researchers' dispositions need to be scrutinizable alongside their methods and findings.

After all the previous sections' discussions, these arguments read like an open door. But consider a notion common to the AI field holding that designing ADM systems to be more humanly explainable comes at the cost of the technology's usefulness.³³⁷ The argument implies the belief that ML systems' version of 'accuracy' cannot exist without a type of complexity that purposefully departs from human understanding capabilities, and that the benefits that are to be had from such systems are a factor of this kind of complexity. Contrarily, Boon for example argues how adhering to a list of 'pragmatic

331 Code, *Epistemic Responsibility*.

332 Williams, *Truth and Truthfulness*, 11.

333 Fricker's own addition to a list of 3, where the 3rd feature of a 'good informer' is 'indicator-properties' Miranda Fricker, 'Rational Authority and Social Power: Towards a Truly Social Epistemology', *Proceedings of the Aristotelian Society, New Series* 98 (1988): 162.

334 Code, *Epistemic Responsibility*, 26–27.

335 After she shifted the Kantian Emphasis away from 'I' earlier, Code, 161–65.

336 Williams, *Truth and Truthfulness*, 143; This quest for truth resembles that of Code's quest for 'wisdom': "the ultimate, possibly unattainable, goal toward which the epistemically responsible strive." Code, *Epistemic Responsibility*, 53–54.

337 David Weinberger, 'Don't Make Artificial Intelligence Artificially Stupid in the Name of Transparency', *Wired*, <https://www.wired.com/story/dont-make-ai-artificially-stupid-in-the-name-of-transparency/>; Against this 'widespread belief that more complex models are more accurate,' see e.g. Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead', *Nature Machine Intelligence* 1, nr. 5 (May 2019): 206–15.

criteria' ensures the human(e) usefulness of all scientific practice *and* its produce. Her arguments for consistency, coherency, simplicity, explanatory power, scope, relevance and intelligibility in their way reunite methods and practitioner disposition. To cite Code, "knowledge claims must be justifiable as much in view of how things are in the world as in view of the human creativity out of which they are formulated."³³⁸

Responsibly sharing knowledge with others is the focus in the second leg of the tandem virtues (sincerity,³³⁹ intellectual honesty, trustworthiness). Virtues, dispositions, and attitudes on this side are about enabling others to trust, rely on, and further develop knowledge claims. Explanation is a salient part of these processes of sharing. What this thesis calls explainers are 'instructors' in Williams's arguments: teachers, judges, scientists—persons in a position of social authority who know that their epistemic claims are taken seriously. Their explanatory activities need to be ruled by the 'critical principle': that the social acceptance of an epistemic power hierarchy needs to be justified on terms that lie outside of the hierarchical relation.³⁴⁰ The central notion here is that doing this responsibly entails explicating relations to sources of quality, trustworthiness etc. that exist outside of the relationship in which the knowledge is shared.

Types and characters of sources to relate to are many and multi-fold: rules & standards, accreditation, other experts, methodological information, oversight bodies, et cetera. The extent to which relations to justificatory sources need explication, and the forms that this should take are what needs serious attention here. In light of the preceding discussion, the lookout would need to be for a possible lack of quality in what is typically shared in—and about—a context, thereby not just reproducing epistemic wrongs but establishing new ones through explanatory acts that disrespects the other epistemic agent. There may be reason to address the who's and hows of rules and standard setting; how diploma systems were designed; why other experts are referred to as such and who invited them into the process; historical, critical methodological

338 Code, *Epistemic Responsibility*, 253.

339 In a related quest, Harry G. Frankfurt characterizes the lack of connection to a concern with truth, the 'indifference to how things really are,' as the essence of his chased object: the definition of "bullshit." As Frankfurt's analysis is much in line with those of Williams c.s., it is a pity he apparently did not read them. He argues that the term "sincerity" is bullshit itself as it signals skeptics' replacement of truth-chasing efforts with a flawed notion of "personal honesty." In fact, Williams describes something like this where he touches upon Rousseau's recourse to "authenticity." Harry D. Frankfurt, *On Bullshit* (Princeton University Press, 2005), 34, 66–67.

340 Williams, *Truth and Truthfulness*, 225–32; Code cites Laurence Bonjour ('Can empirical knowledge have a foundation?', *American Philosophical Quarterly* 15:1 [1978]: 'Cognitive doings are epistemically justified, on this conception, only if and to the extent that they are aimed at this goal—which means roughly that one accepts all and only beliefs which one has good reason to think are true. To accept belief in the absence of such a reason, however appealing or even mandatory such acceptance may be from other standpoints, is to neglect the pursuit of truth; such an acceptance is, one might say, epistemically irresponsible. My contention is that the idea of being epistemically responsible is the core concept of epistemic justification.' Code, *Epistemic Responsibility* .

information. In Williams's terms, the point is to allow others to distinguish 'just' persuasion from 'unjust' coercion of beliefs.³⁴¹

This too may sound rather obvious—and yet, a frequently heard argument against this need with regard to ADM conclusions is that coercion is an accepted, and thus an *acceptable* form of 'knowledge transfer.' A typical example used is that pupils are 'coerced' to accept knowledge claims too. As one particular form of 'denial of crisis,' this rather cynical understanding of education³⁴² ignores how children are not just *told* things, they are *taught* things. Ideally, they learn to appreciate accounts of how things were found out, and by who. They learn how to find and trust sources, to think for themselves-with-others.³⁴³ The fact that we set rules and standards for what proper education entails is an expression of how important we find this. These ideals too need to be non-naïve, since wrongful coercion can and does happen in education.³⁴⁴ In other words, there need to be strategies in place to find and improve upon such wrongs.

3.2.4.2 Doubt, resistance, and a fair distribution of burdens

An example from medical education illustrates the kinds of choices to make in explicating relations to epistemic resources, and how such choices play a role in the cultivation of the kind of epistemic trust a practice aims for, including the kind of 'moral muscles' it wishes to train. Proctor explains that the first convincing research on the (bad) effects of tobacco are from 1930s-40s Germany, and the conclusions, when traced, still hold.³⁴⁵ This research was strongly stimulated by the Nazis, and fell in line with otherwise harmful ideology about pure, healthy and worthy bodies. In

341 Williams, *Truth and Truthfulness*, 251; Another field of research where this need is central is that of unfair commercial practices, see for example Sax, 'Optimization of What?' But these literatures start from a different notion of equality, namely, that which exists between parties to a contract. This is not the type of power imbalance that this thesis focuses on: relations where the social dependency and lesser powers of explainees is a given; More broadly, the online environment is studied through this lens, too Karen Yeung, "'Hypernudge': Big Data as a Mode of Regulation by Design", *Information, Communication & Society* 1, nr. 19 (2016).

342 In Code's words: 'a theory of education in which authority is a matter of impersonal rule-following would have to operate with a truncated conception of the nature of human beings, for whose benefit the system, ostensibly, exists.' Code, *Epistemic Responsibility*, 250; See also Kotzee, who discusses how education's 'epistemic goods' pertain such concepts as true beliefs, justified beliefs, knowledge and understanding, and their teaching should be understood as transformational because it shapes students' "epistemic character" in the process Kotzee, 'Education and Epistemic Injustice', 333.

343 'Teaching refers to explicit presentation and explanation of knowledge, such that it can be the object of debate and discussion' Mireille Hildebrandt, 'Learning as a Machine. Crossovers Between Humans and Machines', *Journal of Learning Analytics* 4, nr. 1 (2017): 6–23.

344 For example, in the US, making the teaching of critical race theory illegal has become a Republican campaign promise David Smith, 'How Did Republicans Turn Critical Race Theory into a Winning Electoral Issue?', *The Guardian*, 3 November 2021, sec. US news, <https://www.theguardian.com/us-news/2021/nov/03/republicans-critical-race-theory-winning-electoral-issue>.

345 Robert N. Proctor, 'The Nazi Campaign against Tobacco', in *Medicine and Medical Ethics in Nazi Germany: Origins, Practices, Legacies*, edited by Francis R. Nicosia and Jonathan Huener, 1st Edition (New York: Berghahn Books, 2002).

other words, fears of ideological taintedness would be well-grounded. But the obvious connotation can also entice to ‘go with’ a less obviously tainted regime. In this case, that of 1950s US about the *harmlessness* of tobacco: claims that turned out to be *deliberately* false. Still that research is much cited as the start of research on tobacco.³⁴⁶

Arguably, teaching *both* research cases would enrich the program in a third, ‘meta-kind-of-way’: by teaching instructors-to-be that it is important to cultivate and maintain a healthy kind of discomfort with regard to their own social-epistemic setting. This is a thesis focused on the responsibility of explainers, and it means to inform their disposition. This further encourages to read Williams’s critical theory *test* a bit differently than he presented it: namely as a possibly useful test for explainers rather than their explainees, or in his terms, for the ‘instructors’ rather than their instructees. What Williams proposes is that the instructed should test a knowledge hierarchy for systemic justness, which should give them an inkling about whether to further investigate truth claims produced in any particular setting. Because the test is not about any other qualities of the truth claims themselves, it is a “weak” test.³⁴⁷ The test entails that the instructed ask a series of questions that should help them to tease out what grounds their belief, trust, or confidence. The answers should lead to justificatory sources that lie outside of the social-hierarchical circles of the instructor. Absent such leads, the knowledge on offer should be labeled as ‘possibly untrue.’ That is where the test ends. Doubt is now instilled in explainees, to return to thesis vocabulary, and ideally they will further investigate: they may still be offered useful knowledge and it would be a shame if they don’t get to benefit ‘simply’ because it is not well justified, even when the explainer has no idea how to go about such justification themselves. In Williams’s terms: the instructors may be of perfectly good faith.

A critical question to ask here is why the lack of justification should *not* lead to the dismissal of a claim. This would leave the burden of ‘proof’ with the claimant—arguably, where it should be. The problem with this is, wrongful claimers may be neither dependent on, nor care about the uptake of their knowledge by explainees in information positions they weren’t interested in in the first place. The Nazi v. ‘Big Tobacco’ ideology example above helps to take this thought further. It illustrates how a knowledge claim from a bad ‘hierarchy’ can still be safe, but also that it would be unfair to place that burden of investigation on the shoulders of a patient. And so for them, explainers with ‘good faith’ need to go the extra mile, taking into account their interests. Williams does acknowledge the substantive investigative burden, but mainly because his is a highly rationalized approach to what ideally happens in more intuitive ways on a daily basis. In other words, that it would be unrealistic, rather than unfair, to go through such tests and investigations each time.³⁴⁸ With the *restorative* aim of epistemic justice in mind, this seems a bit weak, itself. Earlier sections

346 Proctor.

347 Williams, *Truth and Truthfulness*, 229.

348 Williams, 222–24.

described how any ‘instructor’s epistemic wrongs and harms are fed by, and feed into, societal dynamics and shared knowledge spheres. But that does not mean it makes sense to understand epistemic ‘righteousness’ in a similar way. For a fair distribution of burdens, it is not enough to argue that justice-oriented knowledge practices need ‘individual and societal promotion’;³⁴⁹ that individual experience of moral agency lies in the experience of responsibility towards oneself *and* others.³⁵⁰ These things are all true, but they need to be made true, i.e. compulsory for whoever is placed in a position of social and not just epistemic authority *vis-à-vis* explainees. Explainees need to be able to rely on the institutional embedding of all the discussed virtues (both tandem legs), from the level of knowledge practice governance down to that of explanation.

It is of course illusory to assume that all ways and means of cultivating explainers’ personal alertness to possible injustices of their knowledge spheres can be legally enforced; if only because no sphere can be isolated to control how this happens. But the forms of knowledge governance we already have, the fundamental, (supposedly) anti-oppression explanation rules we already have may be usefully further developed. Arguments from the consulted literatures could inform this effort with regard to the behavior that needs to be stimulated. Code argues that those whose epistemic authority is relied on should practice epistemic care, and understand the other’s interests as their own.³⁵¹ She advises to cultivate *discomfort*, described as an aliveness to ‘temptations to endorse and inclinations to reject’ knowledge.³⁵² Williams himself (and in general) adds resistance to laziness and to group pressure.³⁵³ Medina argues how dealing with knowledge comes with a primary responsibility to resist unfair knowledge spheres (“hermeneutical climates”). He describes this resistance as a doubt-imbued, resistance-encouraging variant of what Fricker more idealistically called ‘virtuous listening.’³⁵⁴ Such resistance is also described by Code: a proper pursuit of understanding may entail being “epistemically *irresponsible*, at least in the eyes of [one’s] community”³⁵⁵ —a more elaborate phrasing of ‘group pressure.’

These calls for taking care, for being on the alert for injustices, for the willingness to act in defiance of one’s social-epistemic peers resound in the apologetic reflection of the Childcare Benefits Scandal judges. To reiterate, they had been afraid that if they had acted on the basis of their domain’s perceived injustice, they would have given explainees ‘false hope’ because such judgments risked to be reversed on appeal. In other words, between ‘damned if you do, damned if you don’t,’ they went for don’t

349 Williams, 44–45.

350 Neal, ‘Respect for Human Dignity as “Substantive Basic Norm”’.

351 Code, *Epistemic Responsibility*, 251.

352 Code, 6, 251 (and throughout).

353 And, less relevant, to fantasy, to wish, to believing the more agreeable. Williams, *Truth and Truthfulness*, 133–35.

354 Medina, ‘Varieties of Hermeneutical Injustice’, 48.

355 Code, *Epistemic Responsibility*, 56.

– and ended up ‘letting themselves down.’³⁵⁶ They went for a second ‘don’t’ when they decided to not push protest signals upward in the knowledge hierarchy. A few weeks after the reflections report was published, the Administrative Supreme Court judges themselves issued a report with reflections.³⁵⁷ In it, *they* acknowledged how their judgments had disrespected fundamental values, and explained how they would work to uphold these better in the future. One legal scholar publicly responded that the reflection commanded respect for the judges’ honesty but also gave him stomach pains.³⁵⁸ The judges’ self-realizations were so belated, their suggested improvements so self-evident, that the report begged the question how *this* judiciary would be able to wield its powers responsibly when all other powers fail again in the future.³⁵⁹

The question is pivotal, the central point of similar questions about the fundamental explanation choices of *all* decision makers involved in the victims’ cases. Did their explanation duties not demand that they knew more, understood better, dug deeper, revealed more systematic wrong-doing? When their discomfort was triggered at all, why was it not triggered to the point of disagreement, and of restorative action? If these questions remain unaddressed in how this scandal goes down in history, institutional learnings are not to be expected. After the systematic destruction of the j’s, fundamental legal-philosophical discussions centered the moral responsibility of those who had acted ‘in good faith’ in accordance with Nazi-made laws; the legitimacy *of* these laws, and what constitutes a ‘law’ in the first place.³⁶⁰ In other words, these were discussions about what to do ‘when all other powers fail.’ Constitutions and international treaties were drafted to avoid the need to have such discussions ever again.

Alertness, clearly, needs maintenance. The point is not to draw comparisons between the kinds of harms done in these different times, but to draw attention to the fundamental justification values that would need to be in place to (help to) prevent either. Taking this further, the question is what kinds of harm-establishment the values are imbued with in the first place. When fundamental principles are invoked, the historically grown authority of such terms needs to be qualified in terms of progressive

356 Code, 251.

357 ‘Lessen uit de kinderopvangtoeslagzaken. Reflectierapport van de Afdeling Bestuursrechtspraak van de Raad van State’.

358 Literally, “a knot in the stomach”, a Dutch expression used to describe a somewhat haunting conscientious state. Folkert Jensma, ‘De Raad van State gaat nat, maar is het genoeg?’, *NRC*, 27 November 2021, <https://www.nrc.nl/nieuws/2021/11/27/zelfreflectie-bij-de-raad-van-state-over-toeslagen-hakt-erin-a4066993>.

359 Jensma.

360 A notable, cross-Atlantic discussion is that between legal professors Hart and Fuller, known as the Hart-Fuller debate, H. L. A. Hart, ‘Positivism and the Separation of Law and Morals’, *Harvard Law Review* 71, nr. 4 (1958): 593–629; Lon L. Fuller, ‘Positivism and Fidelity to Law: A Reply to Professor Hart’, *Harvard Law Review* 71, nr. 4 (1958): 630–72.

insight. What to think about the call for an ‘oath of Hippocrates’ for mathematicians?³⁶¹ The original oath did not acknowledge the interests of patients in learning anything about their states. It was to be avoided that patients became *concerned* about their states, and pertinent to command their trust in their doctors’, rather than medicine’s, authority.³⁶² Modernized oaths took centuries to develop, and principled informed consent is even more recent. Most importantly, ethical principles of informed consent were inadequate until practitioners were legally forced to comply. The obligation of taking an oath for physicians by now has been replaced by voluntary ceremonial act in many countries.³⁶³ In light of this brief history already, calling for an oath again with regard to a fast-developing body of knowledge that is not well-enough understood by practitioners and subjects alike seems like a badly informed idea.

3.2.5 Takeaways for ‘explanation’

The preceding sections already allow to understand how explanations (good, bad, or absent) are importantly at play throughout people’s lives: in the context of parenting, education, and in adult interaction, e.g., including consultation, decision making and scientific knowledge practices. Explanation matters for how people learn to function inter-dependently, to recognize the quality of knowledge claims, and the trustworthiness of claimers.³⁶⁴ The following list comprises brief takeaways that shed light on the responsibility of explainers, and the governance of explanation as a practice. These will inform the quest for explanation norms that take the ‘bads’ of knowledge practices into account. In the order in which they were discussed:

- * Structural power inequalities are embedded in liberal, individualistic ideals that abstract from situations and relations of structural oppression. Social institutions in societies where such ideals are strongly represented will have been influenced by them. Such idealizations wrongly assume shared knowledge and mutual understanding between different groups where in fact, the power imbalance also expresses in social-epistemic positions.

361 Ian Sample, ‘Maths and Tech Specialists Need Hippocratic Oath, Says Academic’, *The Guardian*, 16 August 2019, sec. Science, <https://www.theguardian.com/science/2019/aug/16/mathematicians-need-doctor-style-hippocratic-oath-says-academic-hannah-fry>.

362 Needless to say, ‘shared decision making’ was not even a topic, and still had to be argued for in 1984 – as famously done by Jay Katz. Katz, *The Silent World of Doctor and Patient*. 1984.

363 See for example the US <https://www.health.harvard.edu/blog/the-myth-of-the-hippocratic-oath-201511258447>, UK <https://www.bmj.com/content/362/bmj.k3404> and the 2003 Dutch version <https://www.lumc.nl/sub/1060/att/907300228522341.pdf>

364 Defenders of democracy are currently in a panic over fake news and the proliferation of conspiracy theories because the well-functioning of democratic societies depend on these capabilities. With regard to ‘fake news’ cultures, and how technology exacerbates the problems, Gescinska suggests that rather than as ‘post truth,’ our times are perhaps best discussed in terms of their ‘post truthfulness.’ Alicja Gescinska, *Kinderen van Apate, Over leugens en waarachtigheid* (Lemniscaat, 2020).

Explanation ideals established by and upheld in our social institutions, and practiced by explainers in them, deserve to be scrutinized in this light.

* Since all knowledge is a produce of social practice and therefore includes political dimensions, the aim of ‘just,’ responsible knowledge practices lies in avoiding and correcting socially oppressive dynamics and representations. Claims of objectivity as ‘free from human values’ are false. Objectivity in acknowledgment of human values needs to be the aim of knowledge inquiries to avoid relativist exclusions of knowledges and -practitioners: in respect of how all humans are fundamentally equal and inter-dependent, relativism in the sense of ‘what is true for us is not true for them’ should be avoided.

This particular kind of justice needs to inform the responsibility of decision makers/explainers in their dealings with underlying knowledge, and as knowledge practitioners themselves: as conductors of practices wherein ‘conclusions’ about the justness of a decisional process are negotiated.

* Injustices express directly as wrongs, and in harms, when social-epistemic authority is irresponsibly established, informed, and used. Testimonial, hermeneutical, and combined injustices are of direct influence on the social epistemic positions of its victims. This includes their self-understanding, their practical opportunities, and how others understand and respect them. Practices and practitioners that fail to acknowledge this play a role in consolidating such ‘bad goods,’ sustained by the passing of time when this is allowed to let historical sources of social-epistemic oppression sink further away from active societal conscience.

Explainers are in positions of social-epistemic authority. To avoid being instrumental to such wrongful dynamics, these notions should inform (an investigation of) their self-understanding, and their understanding and treatment of others. Both types of understanding stretch to the parties’ respective knowledge spheres, and those which they assume to share (see next point.)

* Persistent and correction-resistant injustices derive authority from, and feed back into, scientific and popular knowledge spheres. Unjust notions at play in populations mean ‘fitting’ wrongful findings are embraced while corrective notions are resisted. ‘Low maintenance’ epistemic oppression leads to excessive burdens of proof on victims’ shoulders.

Institutional explainers are part of (or, ‘produced by’) these societies and power structures. This should lower expectations of corrective labor happening naturally, both for reasons of under-awareness and group pressure. A case for legal rather than (just) ethical governance.

* The chapter described a tandem pair of related values, dispositions, or virtues to align against the described wrongs. The first leg (accuracy/due care/competence) focuses on the *creation of knowledge on responsible terms* and investigative strategies. The second (sincerity, intellectual honesty, trustworthiness) on the *sharing of knowledge in responsible terms*. The point of both is to draw method and disposition together to prevent the making, sharing and perpetuation of oppressive epistemic representations. Necessary traits include a ‘healthy kind of discomfort’ with accepted knowledges and a preparedness to resist one’s group epistemic norms and goods. As part of the justification of knowledge claims, authoritative sources of knowledge need to be related to, their (social-epistemic) trustworthiness explicated. The term ‘epistemic care’ was used to describe how explainers (the term was already used) need to understand the interests of their explainees as their own.

The role authority of explainers as (also) investigators (of decisions, methods, underlying knowledge); as understanders (i.e. explainees, themselves); and as co-creators of knowledge in the form of justifications of how and why decisions were made should be informed by these tandem values. Explainees should be able to count on the institutionalized implementation of this to ensure a fair distribution of investigative burdens.

3.3 Aiming for justice in explanation practice

3.3.1 Meaningful information positions: prerequisite and aim of responsible explanation practices

3.3.1.1 *Recap: the quest for meaningful information positions*

The last part of the chapter introduces additional arguments from the consulted literature—insights that speak to situations of institutionalized explaining quite directly. These will be discussed in sections 3.3.1.2 to 3.3.1.4, after the current section re-introduces what the thesis aims for: an understanding of what meaningful positions are. Section 3.3.3 then models all the chapters findings into a set of ‘duties of care’ for explanation practices, with which an understanding is established of what meaningful positioning, as prescriptive explainer behavior, could look like.

The preceding and following sections sustain a re-idealization of explanation in recognition of non-ideal theory with regard to knowledge practices. The choice for the re-idealization itself was based on clues about apparent non-ideal explanation practices. Assuming these practices express fundamental, anti-oppression explanation ideals, these ideals seem inadequate. Still, reflections from judges involved in a major recent administrative justice scandal in The Netherlands testified to how there is some awareness of what a proper justification practice could have, and should have afforded. In that sense, the scandal, just like the confrontations with explanation challenges that

ADM cases are serving societies with, may be consulted as a usable crisis moment. A quick engagement in Chapter 2 surfaced several clues. The administrative judges expected themselves to engage with rules and fundamental principles of their domain more critically; to not ignore their moral conscience and intuitions when these signal apparent injustices; to follow up on clues that crucial information may be withheld from their scrutiny; to come to a more intimate understanding of the aims, methods and processes of the decisions under scrutiny; to qualify the information positions of their explainees; and to understand the consequences of their judgments, i.e. their defense of the justness of decisions under scrutiny for the larger administrative reality of their explainees.

What was *missing* from these reflections was an awareness of, and engagement with, (possible) racist and discriminatory law, policy, and decision making in the explainees' cases. There were clues that this was happening: individual case clues, aggregate case clues, and political and societal debates (during the many years of the scandal) on the existence of such injustices. These debates also pertained to conceptual societal and political understanding of individual as well as structural, institutionalized racism and discrimination. With that, the reflections, just like the initial judgments, 'reasoned away' the most fundamental dimension of justice that was dis-served. With this, the social-epistemic positions of the cases' explainees was not improved but harmed further. They were also insulted: disrespected as cognitive agents. All in all, the clues inspired to pursue an understanding of explanation as a practice in need of more meaningful explainer information positions, and a practice that should aim to more meaningfully improve explainees' information positions. Only then can established aims of regulated explanation, namely to ameliorate power and information inequalities, be served responsibly.

3.3.1.2 Explanation as an interactive, testimonial practice

To further the investigation of what 'meaningful' explanation is, it was argued that it helps to understand explanation as a knowledge making practice in itself. Knowledge is made about how decisions were made, and as decisions include dealings with knowledge, too, explanation creates knowledge about knowledge. Secondly, it helps to understand explanation as an activity with a social dimension that pertains to the activities of explainers and the interaction between them and their explainees. This is the dimension of conduct; and as there was conduct involved in the decision making under scrutiny, too, norms for the conduct of explainers are also norms about previous conduct. The following sections make clearer how the tandem of values/virtues/dispositions can be engaged in explanation. The values of the first leg, i.e. accuracy, due care, responsibility are *especially* important with regard to the making of knowledge. Sincerity, intellectual honesty, and trustworthiness *especially* see to the behavior while doing this.

These understandings are usefully captured by an understanding of explanation as a relational practice. Relations are inevitably interactive, and so, the outcomes of explanation practices are inevitably collaborative. Even if explainees were to keep silent during the explanation, that does not mean the explainer controls how they are understood. Secondly, it helps to understand explanation as a testimonial practice. Testimony here is understood as a practice of both generating knowledge and defeating false claims to it.³⁶⁵ In our case, an exchange about how a decision was made, or a conclusion was reached, in an institutionalized setting where utterings about this are made to matter. Testimony produces a ‘stamped’ outcome: a claim or declaration that is (scientifically, legally, socially) valid and actionable (accurate, careful, responsible), made by claimers who act truthful (honest, sincere, trustworthy). ‘Actionable’ in the situations the thesis focuses on can mean different things: the decision outcome or recommendation can be accepted or objected to, inform further behavior, or further decisions by other parties and institutions (e.g., eligibility decisions with regard to housing, financial support, education).

Exposure to ‘courtroom drama’ can give the wrong impression that testimony is a one-way process. Someone utters, the other person accepts, or dismisses. But herein already lies a clue about how interaction is inevitably involved. The information positions of explainers and explainees are fed by their respective and shared knowledge spheres. Explainers and explainees both bring their knowledge, experience, and inclinations to trust and distrust, believe or not believe, to the table where claims are *negotiated*—as that is what in fact happens.³⁶⁶ These information positions inform propositions, acceptances, negations, and takeaways. Different parties may take away different things from negotiations, even when false claims are not deliberately made (earlier chapters cited that explanations of how ML predictions are correlative did not prevent explainees to still understand—and use—ML conclusions as causal.) But the previous sections discussed how differences in understanding, trust, and refusal or acceptance may also follow from epistemic injustices, and that outcomes produced this way produce further injustices and direct harms.

Responsible testimonial governance means that practices are designed to prevent the making and uptake of false and unjust claims, including non-deliberate ones. Insights from the consulted literature underscore how the norms that are typically set for who may justify a claim, and how they should go about it, need to be known, understood, and attainable by all parties for the process to be fair. In other words, they should afford ‘due explanation process.’ The need for explainers to obtain an (as) adequate

365 Geuskens, building on Fricker, Williams, and other sources, Geuskens, ‘Epistemic Justice: A Principled Approach to Knowledge Generation and Distribution’, 112; See also Code, *Epistemic Responsibility*, 65, 169–70.

366 I use the term negotiation to signal how the acceptance of a claim can depend on the sincere engagement of the claimer with the potential believer, especially when historical injustices by the institution that the claimer draws authority from are known to them. Grasswick, ‘Epistemic Injustice in Science’, 319–20.

(as possible) understanding of the social-epistemic positions that explainees enter the process with should arguably be part of those norms. The fairness of explanation processes should also not abstract from earlier ‘testimonial moments’ that decision subjects were involved in. Throughout decision processes, subjects (including patients, students) provide information, or information (made) about them is gathered from other sources.³⁶⁷ Norms that govern *these* testimonial moments importantly determine the quality of their participation in decision processes as earlier ‘due process’ norms. And here as well, epistemic injustices (of all kinds) need to be actively avoided.

When the eventual explainers’ understanding of the justness of the bodies of knowledge that ground their decisional process (laws, policy, medical science, assessment methods) is of insufficient quality, if they don’t understand the justness of their methods (manual or digital), or when their understanding of the preceding testimonial moments and *their* justness is insufficient, ‘false claims’ cannot be defeated through the testimonial process. In other words, knowledge can be obscured, or eventually lost through testimony.³⁶⁸ The bad information positions of the judges and their ‘reasoning away’ of racist and discriminatory practices in the Benefits Scandal are a case in point. Bad quality knowledge-about-knowledge gets stamped as ‘justified’ and the product of explanation becomes *worse* knowledge.

With regard to the depth of understanding that explainers should be obliged to attain, Code expects claimers (explainers) to practice due care: to want to know whether, and when, the quality of their knowledge sustains their “right to be heard on it.”³⁶⁹ Williams argues that asserters (explainers) in authoritative positions should act sincere, meaning they should only present beliefs that are appropriate as well as ‘fit’ enough for the context they are uttered in. This means that only appropriate considerations should be made to matter, and that these need to have been accurately investigated.³⁷⁰ This also means explainers need to learn to recognize, and responsibly engage with, ‘tacit’ elements of their cognition like intuition, background, and awareness.³⁷¹ The need for this is also apparent in the direct interaction with explainees, as is discussed next.

367 Public registries, medical files..

368 Geuskens, ‘Epistemic Justice: A Principled Approach to Knowledge Generation and Distribution’, 115.

369 Code, *Epistemic Responsibility*, 92.

370 Williams, *Truth and Truthfulness*, 94; Privacy-theoretical developments of ‘contextual integrity’ also increasingly include more than information types. Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford University Press, 2009).

371 The need for this and arguments against those who are keen to argue how this is ‘unattainable’ is for example usefully discussed in the medical context by Michael Loughlin, ‘Epistemology, Biology and Mysticism: Comments on “Polanyi’s Tacit Knowledge and the Relevance of Epistemology to Clinical Medicine”’, *Journal of Evaluation in Clinical Practice* 16, nr. 2 (2010): 299; Sophie Jacobine van Baalen and Mieke Boon, ‘Evidence-Based Medicine versus Expertise: Knowledge, Skills and Epistemic Actions’, *Knowing and Acting in Medicine*, 2017, 8.

3.3.1.3 Interactional justice demands

The first part of the chapter discussed types of injustice that quite directly affect the interaction between explainers and explainees. Among them, wrongly reduced explainee credibility, a lack of shared epistemic resources, shared oppressive representations. These are obvious obstacles to the kind of ‘interactional justice’ the thesis argues for. This entails respecting explainees as epistemic agents,³⁷² which entails caring for the meaningfulness of their information positions before, during, and after a decisional process. This section adds additional clues about what to take into account in explanation norms so that explainers are adequately instructed. All these considerations bear on the need to strive for *responsible* trust relationships of explainers and explainees; to design a testimonial process that *cares* about trust-worthiness. As Pohlhaus, Jr writes, “[i]f care is the relation that binds moral agents, trust is the relation that binds epistemic agents.”³⁷³ This for example means that explainers need to investigate whether they are trusted or distrusted, and for what reasons.

Even if explainers are trusted, they need to make an explicit effort to earn trust for the right reasons.³⁷⁴ The context of medicine helps to illustrate this. It is a setting wherein explainees are typically unfamiliar with (certain aspects) of the type of knowledge conveyed.³⁷⁵ Proper information sharing as a *sine qua non* condition of a trust relationship is therefore acknowledged in physicians’ professional ethics and training materials. So is ‘explaining per se,’ to familiarize patients with the bodies of knowledge as much as possible, and the need to check patient understanding. Practice however shows how these ideals are hard to attain and need continuous attention and promotion.³⁷⁶ And patients in vulnerable or agitated states can have a hard time understanding and remembering what they are being explained. The Dutch Medical Society therefore asks doctors to encourage their patients to record conversations, and to bring a trusted peer for a second pair of ears.³⁷⁷ But many doctors shy away from such encouragement³⁷⁸ and as we will see in Chapter 5, Dutch legal explanation rules are criticized for their lack of uptake on these particular points.

372 Binns et al, “‘It’s Reducing a Human Being to a Percentage’”; Perceptions of Justice in Algorithmic Decisions’ Based on Colquitt et al’s definition of interactional justice: “the extent to which the affected individual is treated with dignity and respect by the decision-makers.”.

373 Pohlhaus, Jr., ‘Varieties of Epistemic Injustice’, 18.

374 For example by referring to outside sources of legitimacy.

375 Geuskens, ‘Epistemic Justice: A Principled Approach to Knowledge Generation and Distribution’, 118.

376 ‘Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming’ (Utrecht: KNMG, 2004); ZonMw, ‘Achtergrondstudies zelfbeschikking in de zorg’, Evaluatie Regelgeving (Den Haag, 2013) These and other materials are discussed in Chapter 5 that investigates the GP domain. The reader is referred to there.

377 ‘Opnemen van het gesprek’, last consulted 27 January 2021, <https://www.knmg.nl/advies-richtlijnen/dossiers/opnemen-van-het-gesprek.htm>.

378 ‘Opnemen van gesprekken door patienten: Uitkomsten raadpleging KNMG Artsenpanel’ (KNMG,); René Héman, ‘Niet stiekem’, *KNMG - Actualiteit en Opinie* (blog), March 2018.

The need for the described engagements also exists in other decisional contexts that—next to being of instated power imbalance—tend to ‘overpower’ explainees, such as administrative decision making. Applicable to all contexts, Congdon writes how epistemic respect can be expressed by principally *conveying* to prospective knowers (explainees) the acknowledgment of at least a minimal set of epistemic capacities, rights, and responsibilities.³⁷⁹ In our case, ‘rights’ would at least include the right to explanation and what it is for, next to rights that come with the decisional context. Various Data Protection rights have become relevant across contexts, such as the right *not* to disclose certain types of information, or to have wrong information corrected. Teaching graduate students about academic integrity and how to report perceived injustices are an example of explaining responsibilities. Ideally, such ‘acts of recognition’ support a culture of solidarity, inclusivity, and spontaneity.³⁸⁰

Recognition also means that reasons for apparent *distrust* need to be addressed in the interaction. Explainers need to make insightful how they are, or will be (more) trustworthy ‘epistemic partners.’ Distrust thickens when they don’t. As a form of ‘murky,’ this can amount to an epistemic injustice in itself, as it can hold people back from getting what they do in fact need. A case in point is illustrated by the Corona virus vaccination debates that are ongoing at the time of writing. Various groups in Dutch (and other) societies have understandable reasons to distrust novel vaccines, not to mention Public Health policy or their own GPs. But there are also badly informed reasons for distrust, among other things as a result of outright misinformation campaigns. Researchers and GPs have warned vaccination policymakers to engage with distrusting ‘vaccinees’ beforehand; to engage with their grounds for distrust, and further their information positions so that they can make informed decisions.³⁸¹

Put differently: depending on their experiences and situatedness, people may defendably reject knowledge, refuse to provide it, lie, or otherwise not cooperate.³⁸² They may pursue different aims with this behavior. For one, it may be (or seem to be) safer for them. In *Automating Inequality*, Eubanks describes how families in need of social or economic support could be reluctant to supply information that would improve their eligibility scores, because the higher ‘support eligibility’ score with one authority would translate into a higher ‘at risk’ score in another administrative body’s system, and trigger *unwanted* and *unhelpful* interventions from this other body.³⁸³ Like in the previous example, leaving such dynamics unaddressed can make things worse: such families’ ‘responsible distrust’ leaves them without the help they do need, from there to the original situation worsening, and eventually to the type of interventions that they had hoped to avoid.

379 Congdon, ‘What’s Wrong with Epistemic Injustice? Harm, Vice, Objectification, Misrecognition’.

380 Williams, *Truth and Truthfulness*, 45, 127; Code, *Epistemic Responsibility*, 138.

381 Humphreys, Parveen, and Sowemimo, ‘Vaccine Hesitancy’.

382 Medina, ‘Varieties of Hermeneutical Injustice’, 50.

383 Eubanks *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* e.g. 121, 152.

The example however also makes clear that such a family would have been right not to trust a well-meaning case worker / explainer who, unlike them, was unaware of the interplay of the different organizations' scoring systems. This illustrates how it may not just be unwise, but *unjust* to expect decision subjects to cooperate with practices that they know (or think to know) to disrespect them or their community.³⁸⁴ To condemn such families' (re)actions disrespects them as knowers, rights holders, and as subjects of knowledge. Medina argues that extreme socio-epistemic power abuse justifies desperate measures: "to fight epistemically by any means necessary (including the right to lie, to hide, to sabotage, to silence others, etc.)." Mignolo justifies "epistemic in-obedience" on the same grounds, and Pohlhaus, Jr. discusses "strategic refusals to understand."³⁸⁵ These insights advise a need for explainers to engage with 'healthy distrust' in different ways: to check their own critical understanding, and to stimulate and engage with that of their explainees.

3.3.1.4 Beyond statements of reasons: material support for progressive development

To support the progressive development of explanation practices, they need to be study-able. Leaving aside the possibility to study explanation empirically,³⁸⁶ the question is whether the types of documents that are typically produced in explanation practice are adequate study materials. Such records come in the form of judicial and administrative statements of reasons, treatment plans and medical records, eligibility, admission, and grading decisions, et cetera. These documents don't necessarily include a process description—although these may still be created, think of court proceedings. Explainees don't typically participate in drawing them up, although they may have some edit rights.³⁸⁷ They typically summarize the 'end product': the conclusion, decision, and how it was justified. Not necessarily how it was explained: the setting, the interaction, the investigations and exchanges engaged in. Most importantly, they *probably* aren't consciously made up in the understanding of explanation as a knowledge making practice, and with the aim of furthering its quality as such. More research would need to be done on such records to assess their adequacy; for this thesis, the question is whether explanation rules demand that such records are made in the first place. The aim of this chapter is to offer some clues as to why that would be a good idea.

Uses for comprehensive explanation records arguably exist on many levels, or 'knowledge spheres.' For explainers and in their domain, explanation records sustain reflection, training, and assessment. For explainees, their peers, and their communities,

384 Medina, 'Varieties of Hermeneutical Injustice', 50.

385 Medina, 49–50.

386 This is certainly useful but limited in time (past processes can't be studied), accessibility (many situations are confidential), and very labour intensive.

387 That they should be made aware of but frequently aren't: think of medical records, school files, records of judicial proceedings, et cetera.

records that testify to the fairness of underlying rules and knowledge, that tell the story of the explanatory exchange, and that explainees have contributed to may bring (possible) injustices within view that are out of view when unjust decisions are simply stamped as ‘justified.’ This would go some way in preventing that unjustly treated explainees become isolated in their misfortune, unable to explain both to themselves and to others how ‘bad luck’ seems to find them so well. For parties to the explanation relationship together, a document that documents their mutual understanding and (dis/)agreement arguably helps to improve the power and information inequality (an established aim of explanation.) For those involved in explanation research and governance, a growing field in AI-infused times, the usefulness is obvious. More broadly though, making redacted or aggregated explanation records available for outside scrutiny can be part of institutional promotion of social-epistemic values such as “truth, aptness, and understanding.”³⁸⁸ A last example from the Benefits Scandal illustrates this well. At the time of writing, various institutional testimonial processes about the Benefits Scandal have been conducted and are underway (such as Parliamentary inquiries, outsourced consultancy inquiries). These processes are expected to produce actionable knowledge of how and why the scandal happened: actionable for victims (redress and recognition), for (future) law and policy makers, (future) judges and parliament. Among the ‘knowledge claims’ that are being negotiated in these processes, one is of particular relevance, and was cited earlier: whether and how racist and other wrongful discrimination (co-)produced the scandal. Disagreement exists on whether there was discriminatory intent, by who, and towards whom; whether this was rather a ‘by-product’ of laws, policy, or methods, and if that is possible *without* intent; why clues about discrimination were ignored. Whatever comes out of these processes will further inform the understanding of these themes, dynamics and ideologies on the Dutch institutional and societal levels, future lawmakers elected by future voters, and so on. As Geuskens argues: testimonial success on community scale is expressed in terms of gains and losses with regard to its members’ epistemic positions.³⁸⁹

3.3.2 Takeaways for explanation rules

The preceding sections discussed additional arguments to inform the duties of explainers in institutionalized settings, and what a proper explanation practice should afford. These are briefly reiterated here, *not* in the precise order in which they were discussed:

- * The interrelated knowledge and conduct dimensions of explanation (knowledge is made about knowledge, past knowledge-related conduct subsumed in new behavior), were connected to the tandem values. Accuracy, due care, responsibility are

388 cf Pohlhaus who argues that the analysis of epistemic injustices affords an understanding of how our institutional social arrangements can cramp these values Pohlhaus, Jr., ‘Varieties of Epistemic Injustice’, 13.

389 Geuskens, ‘Epistemic Justice: A Principled Approach to Knowledge Generation and Distribution’, 115.

especially important with regard to the making of knowledge. Sincerity, intellectual honesty, and trustworthiness *especially* see to the behavior while doing this.

- * It was reiterated that explainers need to pursue meaningful information positions for themselves, and aim to pursue the same for their explainees. Next to understanding of knowledges, methods etc (see the previous part's 'takeaways') explainers need to learn to recognize and responsibly engage with their own intuitions, background, and awareness. They also need to obtain an (as) adequate (as possible) understanding of the social-epistemic positions of their explainees, and whether they were able to responsibly participate in the decisional process before the 'explanation moment' they are involved in now.
- * Explanation was argued to be an interactive, testimonial practice whose governing rules need to be well understood and expected conduct attainable by all parties. Testimony produces a 'stamped' outcome: a claim or declaration that is considered to be valid and actionable. To make such outcomes useful for parties to the explanation relationship, and enable domain, community, and societal scrutiny and progress with regard to how this is done, a case for co-produced, comprehensive explanation records was made.
- * It was argued that explainers need to explicitly strive to earn trust for the right reasons, whether they are in fact trusted or not. Social and informational overpowering should be actively avoided, and explainees need to be addressed as knowers and rights-holders. Explicating their right to explanation itself needs to be part of this.
- * Reasons for apparent *distrust* need to be addressed to avoid that explainees turn their back on situations that they may benefit from. Refusals to cooperate should not simply be condemned as explainees may have reasons to distrust a practice that explainers are unaware of—but should become aware of. In other words, 'healthy distrust' needs to be upheld and stimulated.

3.3.3 Meaningful information position-ing: in rules

3.3.3.1 Rules say, rules do

Chapter 2 introduced legal explanation rules as norms that typically see to two types of powers: decisional powers and powers of expertise. Decisional powers need justification, and the informational inequality between decision maker and subject needs counterweight. Explanation rules made through democratic lawmaking processes are outcomes of political debates about what needs to be explained about the purpose, process and outcome of decisions and consequential conclusions. Explanations fashioned according to the rules of one domain are used to inform decisions in

another, but overarching principles and fundamental rights and values with regard to ‘explanation’ are valid across domains.

Explanation rules, like all rules, thus have a prescriptive dimension (they oblige to do things) and an expressive dimension (they express a societies’ opinion on a topic.) Both have more explicit and implicit dimensions. For example, in how explanation rules set standards for what needs to be explained, they also set demands for explainability. To some extent they therewith prescribe what kind of decisional processes ‘we allow ourselves to have,’ what kinds are considered to be principally trustworthy for decision subjects (pending explanation.) Furthermore, societal opinions are informed by values, morals, dispositions that typically vary across groups and communities. Rules are therewith also an expression of whose, and which, values are represented in lawmaking processes, and are afterwards enforced (or not) through the legal system.

The same is true for *needs*, in our case social-informational needs. Explanation rules are decisions about what constitutes meaningful information positions: those that explainers and explainees come to the table with, and explainees come out with. They also express assumptions about shared knowledge spheres and mutual understanding, and may set standards for interaction during the explanation process. These more explicitly social aspects usefully relate to a critique that this thesis ‘runs with’: that fundamental, dignitarian values and principles that are called on in discourse about explanation rights sometimes start from a too individualistic picture of human functioning. They therewith ignore how people are inter-dependent nodes in smaller and larger social networks, and how their social-epistemic positions are shaped in good and bad ways through them.

This section proposes a set of explanation norms that explicitly addresses these understandings of explanation rules. The norms are applications of the findings of the previous sections, and re-idealize explanation as a practice of meaningful information positioning.

3.3.3.2 Saying what rules should to: duties of care for regulated explanation practices.

The chapter chooses to model the findings into a set of ‘duties of care.’ A colloquial definition of care duties would describe them as obligations to meet in the execution of particular task. The execution of the care duties themselves is typically qualified by reaching a certain result: if the point is to keep a patient comfortable while awaiting surgery, the fact that they weren’t comfortable proves the duty wasn’t met. As ‘keeping comfortable’ can be done and interpreted in myriad ways and is highly dependent on unforeseeable contexts, demanding a result is easier and arguably safer for the patient than obliging a limited list of actions that may still fail to deliver. But care duties can also describe results that are hard to sustain by evidence. In such cases, care duties

rather describe what needs to be demonstrably aimed for. The care duties of this section include both these characteristics. In the execution of their task ('explaining,') explainers need to demonstrably aim to improve their own and their explainees' information positions. Some ends are explicitly described (such as to make a fuss when they are obstructed to perform a care duty), some are in the form of behavior descriptions to live up to (such as to mitigate social pressure.) Casting these norms as care duties makes them applicable to existing regulated explanation paradigms or those under development, and across decisional domains. The duties' descriptions of what explainers should do 'double' as descriptions of what explanation rules should ask them to do. An analysis on the basis of the modeled duties of explanation care allows to assess the justice aims (expression of values) and justice potential (prescriptive reach and strength) of existing explanation rules. Important to mention is how the thesis, and so the Model, is still predominantly focused on the role of the explainer. It starts from their position, and although it obviously aims to serve the needs of explainees, these can and should be additionally described to suit any explainee group's contextual needs.

The takeaways are distributed over four phases of an imagined 'explanation 'cycle.' The duties first steer the investigative gaze of explainers to their own, domain-relevant social-epistemic information positions; then to that of their explainees. It then describes demands for the interaction that happens between the parties, and promotes the creation of records to learn from for future explanations. Put differently, there are phases of preparation, action, and reflection. In reality, the described behaviors will overlap and intertwine just like the tandem values are ideally simultaneously engaged (with). Presenting them as separate will ensure that none are forgotten, and can be independently assessed to ensure the overall quality of a practice.

The duties themselves are printed in italics. A short commentary is added after each. This is done to familiarize the reader, and the author, with what the duties are for. The duties and their descriptions are reprinted in the domain research chapters that follow. The idea is that the duties will be further developed and improved in practice and (further) research, and this construction will help to keep track of how well the goals express what the thesis meant for them to do.

First duty, or element one: investigating explainer authority

Explainers are obliged to investigate their own social-epistemic positions with regard to their decision-making modalities, and their domain's underlying (input) knowledges in order to assess their role (=explainer) authority: does the explainers' understanding justify their authoritative and trustworthy explainer position? If no (or can't investigate), rebel.

This element obliges that explainers avoid to become an instrument of unjust ('bad', oppressive) knowledge practices, and are able to explain their 'avoidance strategies' to their explainees. To what extent they need to *in fact* explain these strategies is best determined in a decision domain's context. More positively expressed, this element promotes that explainers are able to communicate *how*, and not just *that* they are trustworthy 'knowledge practitioners,' and not just accountable decision makers. The point at this stage is to link the self-reflection of explainers to their position of authority *vis-à-vis* explainees, as part of responsible practice. The need for explainers to rebel exists when explainers feel incapable to do this, for example because they don't have access to justificatory sources or aren't afforded the time, means, or authority to investigate.

Second duty, or element two: engaging with the social-epistemic positions of explainees

Explainers are obliged to investigate the social-epistemic positions of explainees in relation to the decision-making modalities and underlying (input) knowledge at hand; can explainees be expected to responsibly provide (or have provided) the necessary input, and understand the output? If no (or can't investigate), rebel.

This element, like element one, obliges to 'prepare the table' for the negotiation of the how's and why's of decisional outcomes. This time the focus is on how explainees *will be able to* experience a just testimonial process. Explainers need to be able to demonstrate engagement with their explainees social-epistemic situatedness (on individual and group levels) with regard to the larger decisional process and methods: 'the system.' This includes engagement with how a system historically treated explainees as a group and individually. The need to rebel exists when explainers feel their explainees are in no position to participate in the decisional process responsibly.

Third duty, or element three: practicing interactional justice

Explainers are obliged to practice interactional justice, which entails to recognize explainees as knowers and rights-holders. Explainees should be provided information that is proportionate to their pre-investigated and incidental (self-expressed) needs; their knowledge and understanding of relevant, larger & smaller knowledge making processes at hand should be discussed with them with the aim of promoting their responsible (dis)trust; accessible justificatory sources from outside of the authoritative setting need to be pointed out accompanied by instructions on how to follow up on such leads; explainees need to be afforded information about their rights with regard to the explanation and the decision outcome; the possibility of social pressure needs to be mitigated by e.g. allowing to bring allies or make recordings.

The duties of this element describe the interactional dimension and behaviors that need to be given an explicit place in the testimonial process. If any description goes beyond what a process is seen to need, this will need to be justified in the testimonial record. The inclination of lawmakers to treat much practiced (or ‘bulk’) decisional processes as simple, self-evident, ‘routine’ and predictable has led to sub-optimal explanation practices. The implementation of automation in such cases exacerbates the problems while obscuring their origins.

Fourth duty, or element four: creating records

Explainers need to create records of explanation practices. These should be understood as truthful accounts of the testimonial exchange as it was prescribed under element three. Therewith the record should express how all previously described duties were attended to, or provide reasons for when they were not. The records need to be shared with explainees, and made available for outside scrutiny in accordance with rules that govern the decisional domain and relevant privacy and data protection regulation.

These record-related duties are meant to produce more comprehensive accounts than the ‘statements of reasons’ that are typically the outcome of decisional processes. This acknowledges how explanation is a knowledge making practice itself, and therewith a place or conduit of possible oppression. Comprehensive records can sustain progressive development of decision and explanation practices across time and domains.

3.4 Chapter 3 in a nutshell

This chapter ‘re-idealized’ explanation on the premise that our current, fundamental legal explanation ideals and the legal rules based on them do not adequately address knowledge-related harms. It modeled an epistemic justice oriented, epistemic injustice informed set of obligations for explainers to sustain the work of inspecting these fundaments.

An exploration of consulted literature in terms of harms that ensue from the misuse of epistemic authority, in terms of the perpetuation of wrongs in shared knowledge spheres, and of the need for institutional promotion of preventative and corrective labor delivered several takeaways to inform the Model. Additional guidance was derived from a more explicitly explanation practice-oriented engagement with the consulted fields. The chapter argued to question privileged norm setting for explanation; for the ‘always on’ need for explainers to be wary of oppressive representations in the underlying knowledges of their practices and engage anti-oppressive tactics, and for the societal need to make such obligations non-optional.

Several concrete pointers for the right kind of explainer conduct were derived from a set of tandem virtues, or values, described by several authors. The first leg focuses on the creation of knowledge on responsible terms and investigative strategies; the second on the sharing of knowledge in responsible terms. It was argued these descriptions should inform the role authority of explainers as (also) investigators (of decisions, methods, underlying knowledge); as understanders (i.e. explainees, themselves); and as co-creators of knowledge in the form of justifications of how and why decisions were made. Explainees should be able to count on the institutionalized implementation of these values to ensure a fair distribution of investigative burdens.

The explainer-explainee relation was described as an interactive, testimonial practice that needs clear publicly known ‘rules of engagement.’ Explainers need to pursue meaningful information positions for themselves and further those of explainees, and investigate their explainees’ (preceding) possibilities for responsible participation. They need to strive to be trusted for the right reasons only: to explicitly stimulate a healthy kind of distrust. The production of (co-produced) records that serve explainees, allow to check compliance and further the development of explanation governance was argued for.

The Model that accumulates these insights as a set of ‘duties of care’ tends to four phases of an explanation cycle. There are phases of preparation (investigation of explainers own information positions, then those of their explainees), action (explanation interaction between the two parties), and reflection (record creation). In practice the different activities will overlap and intertwine. The care duties are applicable to novel or existing regulated explanation paradigms across decisional domains. They are offered ‘as is’ and to sustain the analysis of existing explanation paradigms to assess the justice aims (expression of societal values) and justice potential (prescriptive reach and strength) of the investigated rules. They include the obligation to ‘rebel’ for explainers if the decisional practices they are employed in don’t allow to meet the obligations.

Care to explain?

4 Meaningful information positioning and legal administrative explanation rules

4.1 Introduction

4.1.1 Function and value of the administrative domain study

The proliferation of novel and complex knowledge and decision making (support) methods and systems has challenged established legal paradigms of individual explanation rights and duties. Challenged them with regard to what they are relied on to afford, but also, as this thesis has thus far argued, with regard to the conceptualization of the pursued ideals. Fundamental values of giving explanation are broadly invoked in contemporary literature on the AI-infused ‘explanation crisis.’ But descriptions of what happens to people in absence of a proper operationalization of these values are of insufficient quality. They tend to ignore that such harms were already a reality for many groups of people in less privileged societal positions. There is reason, in other words, to inspect the fundamental values that are praised, and how they have been codified. If they are set up to fail, re-operationalization will serve only part of humanity again. Such an inspection is all the more necessary since the value of understanding knowledge and decision-making methods itself (rather than just assess their outcomes) is being pulled into doubt, especially by the technological community that builds the systems. Whatever the merits of such arguments are, the assumption that this will be a safe practice for all kinds of decision subjects is naive.

These considerations ground the third research question of the thesis: “how do existing legal rules in two seminal regulated explanation domains promote responsible (non-oppressive, information position improving) explainer behavior?” This chapter reports on the first domain: the rules that see to individual administrative decisions. Several considerations grounded the choice for this domain. To reiterate what was explained in the thesis’s Introduction chapter: to start with, all citizens are subject to administrative decisions. The domain’s explanation rules therewith apply to everyone and their quality is a national concern. Furthermore, administrative bodies are tasked with providing basic support for who needs it (‘the welfare state’) in the constitutional democratic societies the thesis focuses on. This means that the less privileged groups that the thesis is concerned with are represented. An important reason exists in the domain’s standing as ‘foundational’: the rules are considered to be exemplary and are much referred to as illustrations of what fundamental explanation rules should afford. For this reason, the domain’s fundamental principles are much discussed as inspirational with regard to dealing with ADM explanation challenges. Within and ‘outwith’ the domain, the principle of motivation (frequently named together with due diligence, and due process) is nationally and internationally described as a norm that on the one

hand already expresses fundamental explanation needs, on the other as a norm that can and should be further developed to serve decision subjects in AI-infused times.³⁹⁰ A last justification for the choice for this domain lies in how it will be combined with another ‘seminal’ domain. Together, they describe a prototypical rule-based, and a prototypical expertise-based domain. Juxtaposing two very different natured domains provides a broad spectrum of insights – some of which, as it turns out, are quite similar to each other.

The chapter is structured into three parts to provide the insight that is pursued. The first two are descriptive. Part one (section 4.2) charts relevant ‘whats, who’s, and hows’ of administrative decision practices. As simple as that setup sounds, attributed knowledge and decision-making powers are complexly distributed in this domain, and in practice can be hard to identify precisely. Whereas this finding is important to relate and will indeed be elaborated at several points, it also means that a comprehensive domain description was not aspired to. The volume and necessarily legalistic detail would distract³⁹¹ from the challenge at hand, which is to critically analyze the domain from a less traditional perspective. The words ‘functional characterization’ in the ‘what’ section’s title expresses how the section offers a ‘sneak peek’ that allows to gain the necessary (and minimally sufficient) understanding of the quality of the information positions of explainers and their explainees in this domain. For those versed in administrative law, the reading experience may be unsatisfyingly incomplete.

The second part (section 4.3) describes the information positions that explainers are *assumed* to have per their explanation obligations, and the positions they are expected to pursue for their explainees. This part spends considerable time on the codification history of the principle of motivation, since the principle is set up to be further codified again. The third part of the chapter analyses these findings with the use of the modeled duties of explanation care that was developed in the previous chapter. The four elements of the Model are used as categories, to classify the findings under.

The observations that this analysis brings forth have self-standing value in how they provide a necessary evaluation of this explanation paradigm’s justice potential. But they are also functional in how they allow to answer the last research question: “what lessons from the analysis of existing explanation rules can we draw to inform how we deal with ADM explanation regulation?” With this in mind, the last section’s findings are ‘stored,’ to be picked up again in Chapter 6. Together with the observations from the second domain, that chapter extracts lessons that need to be on board in how societies (explainers, researchers, and rule makers) proceed to deal with AI’s explanation challenges.

390 Oswald, ‘Algorithm-assisted decision-making in the public sector’; ‘Ongevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen’; Schlössels, ‘Kroniek beginselen van behoorlijk bestuur 2017’.

391 Indeed, as explained in the Introduction, this was also a reason not to take the chapter research itself further (or deeper if one will.)

4.1.2 Research and reporting choices

The chapter (and the thesis) is written for an interdisciplinary audience, rather than for lawyers, political scientists, or domain practitioners who can be expected to have a professional understanding of Administrative Law and decision making. Mindful of this, some of the research scoping choices that the largesse and legal complexity of the Administrative domain necessitates were hard to make. As was explained in the methods section of the thesis's Introduction, the focus is on how the explanation rules serve decisional practices in which the power and information relations between the State and its citizens arguably require the most 'explanation care': relations in which the information inequality coincides with the social dependency of explainees on the outcome of the decisional process for their well-being and thriving. Either because the latter are in dependent (socioeconomic) states, or because of how the disparate effects of policy works out for them. But this choice still leaves many decisional subdomains within view, each of which arguably needs detailed treatment to render them thoroughly digestible. But by picking one out, some of the functional bafflement that is produced by sketching the larger landscape would necessarily need to be sacrificed. So much was explained before; this section adds some detailed considerations. To start with: the choice was made to start each 'what, who, and how' section with a very general introduction (or: 'deceptively' simple). Relevant depth is added along the way, and two detailed descriptions of what happens in specific subdomains serve to illustrate salient dynamics. Both practices have been around for decades, but are not much referred to in contemporary discussions as prototypical of the 'crisis of legitimacy' the larger domain is said to have entered after the Benefits Scandal.

This crisis itself unfolded in the years of the project and runs as a red thread throughout it. It also presented research challenges. It has amplified, spread, and sharpened existing critique with regard to the lack of trustworthiness and humaneness that meets Dutch citizens in their interactions with administrative bodies.³⁹² Especially those in need of support or in vulnerable states are seen to bear the brunt of this, and intersectional effects apply.³⁹³ Among other actors, the thesis so far cited judges, scholars, reports and the Council of state—but this chapter deals with administrative bodies themselves. To avoid a possibly distracting 'crisis perspective,' this chapter chose to *not* zoom in on the Tax

392 'Ongekend Onrecht: Verslag van de Parlementaire ondervragingscommissie Kinderopvangtoeslag'; Marcel Ten Hooven, 'De verzorgingsstaat is grimmig geworden', *De Groene Amsterdammer*, 21 January, 2021/1 druk, <https://www.groene.nl/artikel/de-verzorgingsstaat-is-grimmig-geworden>; 'De burger kan niet wachten: Jaarverslag van de Nationale ombudsman, de Kinderombudsman en de Veteranenombudsman over 2021' (Nationale Ombudsman, 2022).

393 The fact that discriminatory character of existing legislation (as effectuated in policy) escapes Parliamentary scrutiny in the legislative process, and is also not well addressed by antidiscrimination law has led to a Senate initiative at the time of writing. Parlementaire onderzoekcommissie effectiviteit antidiscriminatiewetgeving, 'Gelijk recht doen: Een parlementair onderzoek naar de mogelijkheden van de wetgever om discriminatie tegen te gaan' (Eerste Kamer der Staten-Generaal.), last consulted 22 September 2022 For an English summary of the Report, see https://www.eerstekamer.nl/intern_stuk/20220607/do_equal_justice_samenvatting/f=/vltsi61u34r1.pdf.

Administration, although the Scandal will still be referenced. Details from different sub domains were selected to provide useful insight into the complex buildup of situations that the contemporary critiques and efforts mean to address (and that are cited as such, see above.) The one case pertains to eligibility decisions for material social (health) care support, the other concerns vehicle registration obligations.³⁹⁴ The chapter will also engage with older critical perspectives on Dutch administrative decision making, as these are arguably prototypical too. The consequences of these choices are that problematic aspects and dynamics of the domain rule the chapter's overtones.³⁹⁵

A further choice to explain here is the limited extent to which administrative decision makers as a type of social actor were engaged with. The chapter chose not to focus on the large bodies of (empirical) literature that exist about their self-understanding and behavior. Although this literature can and should also inform practical governance and interventions,³⁹⁶ the chapter chose to focus on the governance of their information positions and decision-making instructions in Administrative Law. It adds established expectations and critique, based on literature and case illustrations. For the chapter's purposes this is arguably enough. The focus of the eventual analysis is on how they are *instructed to act* in decision and explanation practices, and whether the modeled duties of explanation care are reflected in this. The outcomes of this analysis can usefully inform further empirical research, or action research, in specific administrative domains. The modeled duties of explanation care mean to promote explainers' capabilities in a justice-oriented way—if there are administrative domains for which this is not necessary, that is a good thing. But as was explained, the abundant evidence of the need for improvement is what the chapter went with.

A note on literature ends this section. The thesis chose to focus on two sets of basic, but still fundamental explanation rules; rules that apply to situations common enough for all citizens to encounter. For the administrative domain the choice was to focus on the governance of explanation in primary encounters: first explanations of initial decisions. The consequence for this domain was that case-law, a research source that

394 The chapter chose not to engage with Immigration and Asylum law beside a few references. These domains are certainly worth discussing in terms of Epistemic Justice but they are too specific for the thesis's purposes.

395 In the wake of the scandal various political, judicial, and legislative interventions with regard to standing Administrative traditions and procedures were done and put in motion. This is inevitably consequential for the reporting that the chapter needs to do. The chapter mentions these but will not dwell on the interventions unless they are salient in light of the thesis subject.

396 Much cited in an international context are for example Lipsky's famous analyses of Street Level Bureaucracy[s], With his descriptions of how street-level bureaucrats (SLB's) interpret, work with, and shape rules and policies in the day-to-day governance of their attributed powers, Lipsky paints a world wherein SLB's are shown to also be an effective potential power **against** objectification. Michael Lipsky, *Street Level Bureaucracy: Dilemmas of the Individual in Public Services* (New York: Russel Sage Foundation, 2010); The work propelled much further research, and research about how to conduct street-level research and what can be inferred from it, see for example Peter Hupe en Aurélien Buffat, 'A Public Service Gap: Capturing contexts in a comparative approach of street-level bureaucracy', *Public Management Review* 16, nr. 4 (May 2014): 548–69.

is typically much engaged with in order to explain a legal domain's rules in all their developed richness was not a major source. At (most of) the time of research, the judicial scrutiny of administrative decisions was notoriously 'marginal,' and did not function as a major developmental driver of first instance explanation practices;³⁹⁷ a problem that is not contained to the Netherlands.³⁹⁸ The investigation therefore relied mostly on other sources: (historical) parliamentary papers, municipal documents, Council of State and National Ombudsman (advisory) reports, scholarly literature from (mainly) the legal and socio-legal disciplines, and an authoritative handbook on Administrative Law. In addition, two decades of the rubric 'Chronicles of principles of proper administration' in the Dutch Journal of Administration Law were sourced for relevant case-law developments with regard to the principle of motivation. Some case-law also comes in via the case discussions.

4.2 Individual administrative decision making: a functional characterization

4.2.1 Administrative decisions: "a fact of life" for all Dutch citizens (the 'what')

4.2.1.1 Powers of administrative bodies

This first section very briefly introduces the Dutch administrative decision-making domain: the kinds of knowledge and decision-making powers attributed to administrative bodies, how they are part of 'political climates' but relied on to resist these too; and how their roles, behaviors and accountability are grounded on, and governed by, Administrative Law.

A comprehensive picture of the Dutch 'State' currently comprises of some 1600+ entities;³⁹⁹ from ministries, municipalities, and other well-known public authorities to many other types of institutions, agencies, organizations, and advisory bodies. A substantive amount of them can be qualified as 'administrative bodies,' entities (or parts of entities) attributed with various degrees of policy making and/or executive powers. There are administrative bodies that function as part of e.g., a ministry or municipality (e.g., the Tax Administration is part of the Ministry of Finances, entities of municipalities are responsible for social services), (parts of) private organizations

397 Boudewijn de Waard, 'Proportionality in Dutch administrative law', in *The Judge and the Proportionate Use of Discretion: A Comparative Administrative Law Study*, Routledge research in EU law (New York: Routledge, 2015).

398 See for example Raso, critical of how Canadian legal scholarship should look beyond case law when writing about administrative reasons: 'The decision-making practices of administrative agencies are far greater than those represented in judicial review decisions..' Jennifer Raso, 'Unity in the Eye of the Beholder? Reasons for Decision in Theory and Practice', SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, 1 August 2016), <https://papers.ssrn.com/abstract=2840488>.

399 "Wie vormen de overheid?" <https://www.overheid.nl/wie-vormen-de-overheid>, last consulted October 15, 2022

that attract the status when they execute a specific public task, and bodies that operate independently, only under final ministerial responsibility. These ‘independent administrative bodies’ (are attributed specific powers of policy in expert fields such as public health, road safety, immigration.⁴⁰⁰ Depending on the types of authority that (all) administrative bodies are attributed with, they take part in knowledge and decision making, and norm setting, on different levels. They engage in (generally binding) secondary law making, create (self-binding) policy rules, make guidelines and protocols, and of course, individual decisions. With this, the functioning of administrative bodies is of salient, more and less direct influence on citizens’ social-economic well-being. Citizens are dependent on individual administrative decisions with regard to a wide range of possible actions. They need to apply to administrative bodies for various permits, they are dependent on administrative decisions with regard to subsidies and grants; for social, financial, and other forms of support; for personal information registration, for taxation, and for being recognized as citizens to begin with.⁴⁰¹ They are investigated and punished when they fail to comply with administrative rules. In other words, decisions of administrative bodies are a “fact of life” – not just for all Dutch citizens,⁴⁰² but for whoever resides in the country.

Administrative bodies’ role in a (constitutional) democracy such as that of The Netherlands is typically described in terms of a modern-day version of Montesquieu’s *trias politica*.⁴⁰³ Administrative bodies are part of the executive branches of a legislating State, where both the legislator and the executive are legally ‘kept in check’ by the independent judiciary. The common characterization of Montesquieu’s framework theory as a necessary ‘separation of powers’ is unhelpful here: the point was, and is, to understand the ‘tree’ as a network wherein different types of powers are distributed over, and shared between, different branches, with checks and balances in place to make sure neither trias member works itself up towards totalitarian rule.⁴⁰⁴ For example, the government co-legislates with both chambers of parliament, who are expected to function as ‘contra powers.’ All branches are governed by the same high-level principles of constitutional democracies such as legality, proportionality, and subsidiarity: principles

400 It is precisely the corrosion of (US) Public Agency expertise making as an effect of their digitalized policy outsourcing practices, that Citron and Calo call out as a de-legitimizing problem of Public Agencies. Their article (also) cites how civil servants were dumbfounded in Court when questioned on the hows of individual decision-making practices Calo and Citron, ‘The Automated Administrative State: A Crisis of Legitimacy’.

401 Dutch Administrative Law scholar Damen roughly categorized some 56 types of actions that Dutch citizens need the Administration for, either with regard to permission, non-obstruction, or support. Leo Damen, ‘De autonome Awbmen?’, *Ars Aequi* 66, nr. 07–08 (2017) To be sure, as was said, the administrative domain of immigration law was mostly left out of scope in this chapter.

402 K. J. de Graaf et al, ed., *Quality of Decision-Making in Public Law: Studies in Administrative Decision-Making in the Netherlands* (Groningen: Europa Law Pub, 2007), 3.

403 Baron de Montesquieu, *L’esprit des Loïs*, 1748.

404 Montesquie Instituut / kenniscentrum parlementaire democratie, ‘Trias politica: machtscheiding en machtspreiding’, last consulted 11 November 2022, https://www.montesquieu-instituut.nl/id/vhnnmt7lidzx/trias_politica_machtscheiding_en.

with the salient function to keep a check on the “contemporary [political] delusions” that democracies are typically sensitive to.⁴⁰⁵

To further introduce an important dimension of administrative knowledge making and norm setting: powers to create ‘generally binding regulations’ are typically attributed in primary laws, which themselves are enacted by the democratically chosen legislator: Government and Parliament. The specificity of the wording with which these powers are established varies, with which administrative bodies are left with less or more, and as we will see sometimes very large discretionary space with which to interpret and explicate the underlying law. And like the generally attributed (per the General Administrative Law Act, introduced later), self-binding policy rule and guideline making powers, their form is also ‘free’ to the extent that whether something amounts to a rule is frequently established in court.⁴⁰⁶ Democratic process requirements for primary lawmaking such as parliamentary control and Council of State advisory reports do not apply to these rule-making processes, nor are the norms that they produce subject to judicial scrutiny in The Netherlands, save exceptions – although decisions based upon them are. This already makes the norm setting influence of the ‘executive branch’ considerable. When one takes their close collaborations with government(s) into consideration, and the closeness of the government and parliament,⁴⁰⁷ questions can be (and are) raised about the contemporary adequacy of checks and balances in aforementioned *trias* notions.⁴⁰⁸

Administrative bodies, in turn, can delegate executive powers to other bodies, and to private entities. As will be further discussed in the section that introduces Administrative Law (4.2.1.3), the specific powers of different types of administrative bodies, or even their identity *as* such a body aren’t necessarily apparent and can be hard to establish even in court. This is especially (but not only) unclear in cases where the State makes use of persons or (parts of) entities, including private ones for the realization of public tasks and goals: the so-called ‘B’ bodies that are not explicitly established like ‘A’ bodies are.⁴⁰⁹ It is an established judicial challenge to ensure that the State does not skirt its constitutional responsibilities through such constructions.⁴¹⁰

405 Ten Hooven cites Professor of Constitutional Democracy Dorien Pessers on the failure to heed these principles by all *trias* members in the Childcare Benefits Scandal. Ten Hooven, ‘De verzorgingsstaat is grimmig geworden’.

406 Tollenaar writes how this also expresses the purposeful reliance on the ‘absorbing capability’ of Administrative Law. Tollenaar, ‘Bestuursrechtelijke normering en “big data”’, 134.

407 Alex Brenninkmeijer, ‘Welke lessen zijn te trekken uit de kinderopvangoeslagaffaire en de problemen bij uitvoeringsorganisaties?’, in *Grensoverstijgende rechtsbeoefening: Liber amicorum Jan Jans*, edited by K. J. de Graaf e.a. (Zutphen: Paris, 2021), 200.

408 There are more reasons that parliamentary control is considered to be limited, but these discussions go beyond the scope of the chapter. Brenninkmeijer, 205; W.J.M. Voermans, ‘Besturen met regels, volgens de regels’, in *Algemene regels in het bestuursrecht*, Preadviezen Vereniging voor Bestuursrecht, VAR-reeks 158 (Boom Juridisch, 2017), 66.

409 Article 1.1, under ‘a’ and ‘b’, Awb.

410 Willemien den Ouden en Heleen van Amerongen, ‘Het bestuursorgaan-begrip voorbij?’, in *De conclusie voorbij. Liber amicorum aangeboden aan Jaap Polak*, edited by M Bosma e.a. (Ars Aequi, 2017), 149.

Administrative bodies are also referred to as bureaucratic bodies. Modern bureaucracies are relied on as a stable force against any elected Government's willful (attempts at) power abuse.⁴¹¹ And in how they are the 'record keepers' of consecutive governments or 'administrations,' they are instrumental to a steady, continuous form of parliamentary and judicial control. As the Dutch Council of Public Administration puts it, "the democratic, constitutional state in part finds its legitimacy in its bureaucracy."⁴¹² But historically comprehensive descriptions of 'bureaucracy' need to make room for much less positive characterizations. As instruments of government power, bureaucracies were developed through histories and ideologies of colonialism and exploitative capitalism: bureaucracy enabled and effectuated human trade and trafficking, human and geographical exploitation.⁴¹³ In the Netherlands, the horrors of these colonial histories are a lesser part of educational curricula, less commonly known and less referred to than the bureaucratic work forces of Nazi Germany are. The latter are typically referred to (also on European legislative levels, as was introduced earlier) as the worst that 'bureaucratic armies' are capable of. McQuillan, topically, connects bureaucratic harm and AI-as-knowledge-making harm, and cites Arendt's concerns about humans becoming slaves "not so much of our machines as of our know-how."⁴¹⁴ It is not to downplay this status that it needs mentioning that a backlog of necessary studies of Dutch administrations' role in enabling colonial horrors, which took place earlier, is being engaged with in recent years. These enable to know, to teach, and to trace and deal with consequential effects on contemporary populations in former occupied lands as well as in The Netherlands.⁴¹⁵ Indeed, books and (other) studies that make specific collaborations of Dutch bureaucracy with Nazi occupancy insightful are similarly conducted in this day and age more than before.⁴¹⁶ The well-

411 René ten Bos, *Bureaucratie is een inktvis* (Amsterdam: Boom Uitgevers, 2015), 14.

412 'Hoe hoort het eigenlijk? Passend contact tussen overheid en burger' (Raad voor het openbaar bestuur, June 2014), 7.

413 The study of bureaucracies in light of these legacies will always remain urgent if we are serious about understanding the powerful tool of oppression that bureaucracies can be. See e.g. Muhammad Azfar Nisar en Ayesha Masood, 'Bureaucracy and the Other: A Systematic Review of Postcolonial Scholarship in Public Administration', SSRN Scholarly Paper (Rochester, NY, 14 July 2021), <https://doi.org/10.2139/ssrn.3886409>; See also Prince on the colonial roots of 'statistics' in particular Russell Prince, 'The Geography of Statistics: Social Statistics from Moral Science to Big Data', *Progress in Human Geography* 44, nr. 6 (15 September 2019): 7,8.

414 McQuillan, *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*, 62.

415 See for example Pepijn Brandon et al, red., *De slavernij in Oost en West: het Amsterdam onderzoek*, 2021 Similar studies were recently done about Rotterdam, resulting in three books: 'Het koloniale verleden van Rotterdam,' Edited by Gert Oostindie; 'Rotterdam in slavernij,' authored by Alex van Stipriaan; 'Rotterdam, een postkoloniale stad in beweging,' authored by Francio Guadeloupe.

416 Rood, Lentz. *De man achter het Persoonsbewijs*; Bianca Stigter, *Atlas van een bezette stad* (Atlas Contact, 2019); Rob Bakker, *Boekhouders van de Holocaust: Nederlandse ambtenaren en de collaboratie*, Holocaust bibliotheek (Hilversum: uitgeverij Verbum, 2020); Peter Romijn, *Burgemeesters in oorlogstijd: besturen tijdens de Duitse bezetting* (Amsterdam: Balans, 2006); Coen Hilbrink, *'In het belang van het Nederlandse volk...': over de medewerking van de ambtelijke wereld aan de Duitse bezettingspolitiek 1940-1945* ('s-Gravenhage: Sdu Uitgeverij Koninginnegracht, 1995).

known Nazi reference, in other words, has arguably been underinformed. This should not surprise: acknowledgement of racism and other discriminatory tendencies on State levels is hard-won in The Netherlands. The National Ombudsman, tasked with keeping ‘the executive’ in check, has warned about it more than once in the years preceding the Benefits Scandal revelations.⁴¹⁷

4.2.1.2 *Administrative bodies and political climates: realistic expectations*

And so, it is important that contemporary descriptions of administrative bodies do not abstract from how they are inevitably part of, and sensitive to, political climates—at least not when a realistic expectation of bureaucratic resistance to oppression is pursued. And realistic expectations certainly need to be part of any re-idealization of justificatory practices, in this case of administrative decision making.

At the time of writing, the Childcare Benefits Scandal produces much scholarly and also political reflections about the origins of wrongful treatment by public authorities and their administrative bodies.⁴¹⁸ Political reflections are received as belated, argues the former National Ombudsman. He voiced his indignation about the Government’s purported surprise when the scope and depth of the wrongs and harms of the Benefits Scandal eventually ‘presented itself’ to them. The necessary information on what was going down had been available to them for many years, and many problems had been reported and presented to them in various Ombudsman reports and through their own administrative information channels.⁴¹⁹ In other words, they knew, and by not acting on that information, they effectively condoned the Tax Administration’s self-created policies. This behavior in turn is seen to have influenced the Administrative Judiciary, who assumed the Tax Administration (and other administrative bodies) effectuated the Lawmaker’s harsh will where this was in fact not made explicit to such extent in the primary law—raising questions about their own political neutrality.⁴²⁰

But the Scandal is not the first example of how harsh political climates express in administrative policy making; it is rather one in a line of many. The tendency to ‘crack down’ on citizens in dependent positions is described as part of broader European

417 Brenninkmeijer, ‘Welke lessen zijn te trekken uit de kinderopvangoeslagaffaire en de problemen bij uitvoeringsorganisaties?’, 202–3.

418 “The infringement on good faith by all trias members is as incomprehensible as it is unforgivable,” Ten Hooven, ‘De verzorgingsstaat is grimmig geworden’, 17; Lukas van den Berge, ‘Bestuursrecht na de toeslagenaffaire: hoe nu verder? Over het rechtskarakter van het bestuursrecht’, *Ars Aequi* 2021 (November 2021): 987.

419 One of many interviews where he voices his public surprise, Lammers, ‘Oud-ombudsman Alex Brenninkmeijer ziet in de toeslagenaffaire geen bedrijfsongeluk maar een falend systeem’.

420 Bert Marseille en Alex Brenninkmeijer, ‘Een dialoog met de Raad van State na de toeslagenaffaire’, *NJB* 2021, nr. 8 (26 February 2021): 608; Leonard Besselink, ‘De Afdeling Bestuursrechtspraak en de rechtsstatelijke crisis van de Toeslagenaffaire’, *NJB* 2021, nr. 3; Leo Damen, ‘Ik was het niet, ik was het niet, het was de wetgever!’, 4.

political developments with regard to the transformation of ‘the social welfare state.’⁴²¹ These were built up after the Second World War and held in generally high regard for the first decades. From the late 1980s financial crisis onward, the political climate became increasingly intolerant of (especially) the growing unemployed workforce, instigating moves towards (especially) welfare fraud detection.⁴²² Over time, the aim of cutting welfare expenses and a growing distrust of how citizens made use of all kinds of support rights increasingly expressed itself in primary and secondary laws and policy, but again, not always explicitly.⁴²³ Along the way, the ‘othering’ of support needing subjects became increasingly tinted. As Achbab described, immigration laws harshened simultaneously with those on social security, which trickled over in policy developments of the less citizen-friendly laws.⁴²⁴ He argues that an image of ‘parasitical,’ undeserving ‘others’ gradually tainted both public policy domains, adding what are now acknowledged to be racist and discriminatory dimensions to what were already ‘harsh’ policy translations.⁴²⁵

A notorious set of Dutch welfare laws that effectuated these harsher political climates were grounded on notions of citizen self-sufficiency and bureaucratic capabilities; ignoring social research that had repeatedly revealed a different picture (and citing no evidence to ground the Governmental view).⁴²⁶ The laws especially estranged from realistic capabilities and needs of citizens in vulnerable states and situations.⁴²⁷ The overhaul meant to “re-calibrate” the relations of citizens and Government.⁴²⁸ Support affordances were severely restricted, and meant to be made available only when

421 Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights in the case of NJCM /De Staat der Nederlanden (C/09/550982/HA ZA 18/388), ‘Brief by the United Nations Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM C.s./De Staat der Nederlanden (SyRI) before the District Court of The Hague (case number: C/09/550982/HA ZA 18/388)’, 26 September 2019, <https://www.ohchr.org/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf>.

422 J. Oerlemans en Y.E. Schuurmans, ‘Internetonderzoek door bestuursorganen’, *Nederlands Juristenblad* 2019, nr. 20 (2019); F.M. Noorman, ‘Quality and Administration of the Dutch Social Security System: An Impression’, in *Quality of Decision-Making in Public Law: Studies in Administrative Decision-Making in the Netherlands*, edited by K. J. de Graaf et al. (Europa Law Publishing, 2007), 10.

423 De Nationale Ombudsman, ‘Burgerperspectief: een manier van kijken’, Jaarverslag 2015 (Nationale Ombudsman, 2015).

424 Achbab, ‘De Toeslagenaffaire is ontstaan uit institutioneel racisme’.

425 Achbab; See also the recent Senate initiative that was cited earlier Parlementaire onderzoekscommissie effectiviteit antidiscriminatiewetgeving, ‘Gelijk recht doen: Een parlementair onderzoek naar de mogelijkheden van de wetgever om discriminatie tegen te gaan’.

426 A. Tollenaar, ‘Empathie in het sociaal domein’, *RegelMaat* 33, nr. 3 (May 2018): 138; ‘Incomprehensible Government’, Summary of the 2012 Annual Report of the National Ombudsman of The Netherlands’ (Nationale Ombudsman, 2012).

427 J.B.J.M Ten Berge, “De mythische burger: strategische mensbeelden in het bestuursrecht,” *NTB* 2011, no. 22 (n.d.), accessed April 1, 2019.

428 ‘Terug aan tafel, samen de klacht oplossen: Behandeling klachten over zorg, jeugdhulp en begeleiding naar werk’ (Nationale Ombudsman, Maart 2017).

evidence of the inadequacy of informal care circles was established. Citizens were assumed to be ‘calculative’ and needed to be made clear that they were expected to uphold ‘their end of the bargain.’⁴²⁹ Included in the new regime were rules for ‘social returns’ (e.g. obligatory community services), stringent anti-fraud regimes, including punishment for accidental on-compliance and what was seen as unwillingness to find paid work, including ‘unprofessional dressing habits’ or in Tollenaar’s words “clothes that frighten potential employers.”⁴³⁰ Administrative bodies were attributed with a very broad discretionary space in terms of policy translation and execution, and the administrative legal paradigm was relied on in terms of due diligence, explanation and justification.⁴³¹

Critique on how the Government vision was unfounded and unjust was on the table from the start. It increased over time as Ombudsman institutes, the Government Scientific Council and other authoritative bodies investigated the new regimes.⁴³² A ‘re-calibration’ had indeed established; but rather than the ‘horizontal’ ideals that the Government had professed to foresee, the strict compliance regime (had) led to a highly paternalistic regime and mutual distrust.⁴³³ In 2019, a former Minister who was co-responsible for implementing the laws conceded that the citizen model was incorrect.⁴³⁴ A political party member, more elaborately, argued that they had put large numbers of already vulnerable citizens in harm’s way. By starting from assumptions of responsibility and blameworthiness rather than social-economic reality and bad luck, the regimes (further) disabled citizens whose enablement should have been the aim.⁴³⁵ The case discussion in section 4.2.3.5 adds details about one particular type of support eligibility policy that is of particular interest with regard to (the governance of) ‘information positioning’ in the domain. It reveals a different kind of resistance: to political and judicial interference with critiqued, self-made policy.

429 Tollenaar, ‘Empathie in het sociaal domein’, 136.

430 Albertjan Tollenaar, ‘Maintaining Administrative Justice in the Dutch Regulatory Welfare State’, *University of Groningen Faculty of Law Research Paper* 2016, nr. 24 : 6; Tollenaar, ‘Empathie in het sociaal domein’, 136.

431 Tweede Kamer, vergaderjaar 2016–2017, 34 477, nr. 10

432 Linhorst, cited by Schlössels R.J.N. Schlössels, ‘Kroniek beginselen van behoorlijk bestuur 2018’, *Nederlands Tijdschrift voor Bestuursrecht* 2018, nr. 9 (21 October 2018); ‘Terug aan tafel, samen de klacht oplossen: Behandeling klachten over zorg, jeugdhulp en begeleiding naar werk’.

433 R.J.N. Schlössels, ‘Kroniek beginselen van behoorlijk bestuur 2016’, *Nederlands Tijdschrift voor Bestuursrecht* 2016 (2016) para 1; Damen, ‘De autonome Awbmen?’

434 Remarks made at the time of her inauguration as professor at the University of Leiden, where she will hold a chair on ‘Science, policy and societal impact with an emphasis on care.’ Kammer en van Lonkhuyzen, ‘Oud-minister Bussemaker gelooft niet meer in de participatiemaatschappij’.

435 MP Gijs van Dijk in Stellinga en De Koning, ‘PvdA vindt eigen Participatiewet mislukt’.

4.2.1.3 *How all this is legal: the law that regulates powers of, for, and against the administration.*

The attribution of rule and policy making and executive powers, the discretionary decision-making space that administrative bodies have, and how they can be held accountable for their use of it, is mostly governed by the administrative legal paradigm. The primary laws referred to earlier, as laws *for* the administration, serve as a basis of administrative bodies' powers to legislate details, design policy, and make individual decisions. These laws themselves include provisions that serve to protect citizens *against* administrative abuse of power, but fundamental protections are also present in the General Administrative Law Act (Awb) that governs all actions of the State and its administrative bodies. As will be discussed further on, the Benefits Scandal surfaced fundamental differences of opinion in scholarly and judicial interpretations with regard to the hierarchy of protections in specific laws and those in the Awb. At the time of writing, a novel legal proposal aims to strengthen the Awb as "the central, standard setting law" with regard to rights-relevant relations of citizens and State, and of citizens *against* the State.⁴³⁶ It mentions how (in the wake of the Benefits Scandal) even a strengthened Awb should however not be relied on as a 'catch all' for "uncarefully considered, too harsh, or unenforceable laws and rules."⁴³⁷

Other regimes can also apply to actions of administrative bodies. The State can be held to Contract Law when it enters into contractual relations with private entities, and Tort Law applies in instances of apparent misbehavior by public authorities and their administrative bodies. It works the other way around as well: administrative rules are made to apply to semi/private entities with regard to their performance of a public function, such as to a garagist who performs an MOT (mandatory periodical technical checkup for motor vehicles, 'APK' in Dutch). The latter example is straightforward. In reality, also with reference to earlier, the paradigm is complex.⁴³⁸ It can be challenging for citizens and professionals (lawyers, civil servants, judges) alike to figure out which public entity has what status, at what time, relative to what function; to understand precisely what primary legal authority a (national or local) administrative body grounds their rule- and policy making powers on, and which is the legal extent of

436 Ministerie van Algemene Zaken, 'Memorie van Toelichting wetsvoorstel Wet versterking waarborgfunctie Awb (pre-consultatieversie 18 January 2023)' (Ministerie van Algemene Zaken, 18 January 2023), section 2.1.

437 The proposal came out too late to incorporate into the text. The text was scanned, and some useful footnote-level additions were made. The proposal acknowledges the lack of protective force of the administrative legal regime and especially that of the Awb. Recommendations also see to 'motivation'. Although these parts do acknowledge some of the problems this thesis signals, its proposed solutions are disappointing and mainly argue for understandable wording (paragraph 3.2.) Ministerie van Algemene Zaken, 'Memorie van Toelichting wetsvoorstel Wet versterking waarborgfunctie Awb (pre-consultatieversie 18 January 2023)'.

438 A.T Marseille et al, red., *25 jaar Awb: in eenheid en verscheidenheid* (Deventer: Wolters Kluwer, 2019), XIV.

their discretionary decisional space—in general as well as in particular instances.⁴³⁹ For decision subjects, the most direct relevance of this situation is that the extent to which administrative law governs any entities' behaviour depends on the authority they are attributed with. When this is unclear, citizens' access to justice through e.g. administrative due diligence and legal procedure suffers accordingly, forcing them to turn to the more expensive (and financially risky in case they lose) civil law regime.⁴⁴⁰ To foreshadow here is how a second particularity of the Awb regime can do that, too: the law severely restricts what counts as an administrative decision, leaving citizens no recourse than the private legal regime in case they want to object to administrative treatment that does not qualify as such. Section 4.3.1.1 deals with this specifically.

The domain's complex legalistic detail is criticized and cherished at the same time: criticized for establishing a sphere of formality and inflexibility that is not conducive to the flexibility that 'dealing with societal 'reality' requires,⁴⁴¹ cherished because a strong legal regime helps administrative bodies to provide consistency and legal certainty.⁴⁴² Section 4.2 describes 'the how' of individual decision making, and adds details about administrative rules for making individual decisions.

General and specific legal *principles* are part of the described legal paradigm, and govern it to varying extent.⁴⁴³ They serve the explication, justification and interpretation of specific legal provisions that individual decisions are based on, and can be called upon to go *against* a legal rule "under exceptional circumstances."⁴⁴⁴ Schlössels and Zijlstra describe the originally Anglo-Saxon principle of 'fair play' as a kind of mother principle, a broadly shared 'notion of justice' that governs (or should govern) all interactions between public institutions and citizens. Among other things it demands openness, honesty, to adequately and 'wholly' inform citizens, and

439 See e.g. R.J.N. Schlössels, 'Discretionaire dogmatiek... anders de Afdeling bestuursrechtspraak?', *Nederlands Tijdschrift voor Bestuursrecht* 2018, nr. 52 (2018); To illustrate: cases exist with regard to what extent the Tax Agency is an administrative body as a whole, and / or that this status should be attributed to the Inspector; which of the Agency's organizational elements are administratively responsible in Benefits Scandal cases ECLI:NL:HR:2021:995, Hoge Raad, 19/03033, No. ECLI:NL:HR:2021:995 (Hoge Raad 25 June 2021).

440 den Ouden en van Amerongen, 'Het bestuursorgaan-begrip voorbij?', 139, 147.

441 Lukas van den Berge, 'Bestuursrecht Tussen Autonomie en Verhouding: Naar een Relationeel Bestuursrecht' (Utrecht University, 2016), 4; A.C.M. Meuwese, 'Grip op normstelling in het datatijdperk', in *Algemene regels in het bestuursrecht*, Preadviezen Vereniging voor Bestuursrecht, VAR-reeks 158 (Boom Juridisch, 2017), 158.

442 Schlössels, 'Kroniek beginselen van behoorlijk bestuur 2018' 1.1.

443 The situation in The Netherlands is somewhat complex due to the prohibition of judicial constitutional checks of primary legal rules, which stretches to the principles that rule these rules. But, rules of international and treaty law that apply in The Netherlands prevail the constitution, and so these can (and are) be called upon - including principles recognized in these. Widdershoven, 'Een ervaring als staatsraad-generaal: op zoek naar een rechtsbeginsel', 95.

444 Widdershoven, 91.

to treat them properly.⁴⁴⁵ These elements are recognizable in descriptions of other principles, such as that of ‘equal treatment’ which itself belongs to the fundamental (or ‘constitutional’) level that rules the administrative space as well. Principles of ‘good governance’ apply through how they govern the behavior of national governments,⁴⁴⁶ and principles like the Dutch ‘General Principles of Proper Administration’ (GPPA, ABBB in Dutch) govern specific governmental tasks and institutions. In Damen’s words, “specific principles that govern the administrative relations between citizens and the administration, the interpretation of which can be grounded on general legal principles.”⁴⁴⁷ New principles also develop, or are adopted from other ruled domains and applied.⁴⁴⁸ A last type of principle to mention are ‘unwritten’ principles of law and justice. In 1976, the fore-runner of the Awb codified what had been standing, but debated practice in previous decades: that administrative judges could ground their decisions on *unwritten rules of law* and *general legal principles* that ‘live in the general constitutional conscience.’⁴⁴⁹ Writing about (administrative) judicial acknowledgement or ‘discovery’ of such principles, Attorney-General Widdershoven explains how this entails the recognition of certain key features: a measure of vagueness, abstraction and flexibility, a certain weight, and the extent to which they (can or do) qualify as a legal standard.⁴⁵⁰

In how (written / unwritten) principles refer to what is ‘alive’ in societies, they have a role to play in furthering legal development by infusing it with progressive insight.⁴⁵¹ Principles themselves can also benefit from legislation however. As Gerards argues, it can clarify them, create public awareness, and force compliance.⁴⁵² Principles that are already codified can be explicated through such efforts too. The administrative principle of motivation that is discussed at length later on stands to benefit from explication, the

445 R.J.N. Schlössels en S.E. Zijlstra, *Bestuursrecht in de sociale rechtsstaat 1*, 6e druk (Wolters Kluwer, 2010) section 8.3.1, under 13.

446 ‘Considering that good administration is an aspect of good governance ..’ Council of Europe, *The Administration and You, A Handbook: Principles of administrative law concerning relations between individuals and public authorities* (Council of Europe Publishing, 2018).

447 Cited by Van Eck, van Eck, ‘Geautomatiseerde ketenbesluiten & rechtsbescherming: Een onderzoek naar de praktijk van geautomatiseerde ketenbesluiten over een financieel belang in relatie tot rechtsbescherming.’

448 Or added to: new principles have been ‘discovered’ before, such as the precautionary principle. It originated in environmental law environments, and holds that a “proactive stance” should inform an Administration’s dealings with uncertainties. It is now well known outside of it, for example in the GDPR, where it is codified as having a role in dealing with uncertain effects of new technologies. J. H. Gerards, ‘Meer rechtsbeginselen in de Awb? Gezichtspunten voor toekomstige codificatie’, in *15 jaar Awb*, edited by Tom Barkhuysen, Willemien den Ouden, en J.E.M. Polak (Boom Juridische Uitgevers, 2010).

449 Gerards, 787.

450 Widdershoven, ‘Een ervaring als staatsraad-generaal: op zoek naar een rechtsbeginsel’, 88.

451 Widdershoven, 88 section 2.2.

452 Something that was touched upon earlier in the discussion of explanation rules’ expressive function. Gerards, ‘Meer rechtsbeginselen in de Awb?’.

thesis will argue. As was mentioned in Chapter 2, Widdershoven writes how principles aim to answer the ‘why’ behind the ‘what’ that legal rules explain.⁴⁵³ Gerards also discusses possible adverse effects of codification. Codification can become an obstacle for further development of a principle, and lawmakers may (inadvertently) downplay the importance of not or less elaborately codified principles. With this in mind, she argues to govern the codification of principles carefully.⁴⁵⁴ Schlössels interprets her arguments as a warning to engage in precise, rather than abstract codifications.⁴⁵⁵ These discussions have become very pertinent at the time of writing. The Scandal pulled the strength of justice serving principles into doubt, and discussions about more, or different codifications have developed into discussions about the actual and ideal character of Dutch Administrative Law as a whole. The earlier mentioned proposal aims to strengthen the codification of the General Principles of Proper Administration in light of their inadequate force in the Benefits Scandal years.

4.2.1.4 *Did Dutch Administrative Law’s main injustice valve dysfunction? The ‘hardship clause’ discussion.*

In Dutch Administrative Law, the principle of proportionality was codified as a norm that sees to protect citizens from hardship in the form of disproportionate adverse effects that the application of a legal provision would amount to in their situation. It is codified in the Awb’s article 3:4, under 2: “[t]he negative consequences of a decision shall not be disproportionate relative to the objectives that are pursued by the decision.”⁴⁵⁶ The provision is preceded by 3:4’s first paragraph, which instructs administrative bodies to balance all directly relevant interests insofar as their discretionary space, or any specific higher order rule, does not restrict to do so.

A full discussion of the history of this codification goes beyond the scope (and purpose) of this chapter, but some background helps to understand the clause a bit better, which helps understand the discussions *about* the clause after the Benefits Scandal. De Waard writes how the 1994 Awb meant to codify a developed judicial tradition of (only) marginally testing for manifest State abuse of power (the ‘arbitrariness’ test) on the one hand, and to create an instructive norm for administrative conduct on the other.⁴⁵⁷ Using the term ‘proportionality’ in combination with the double negative (*not dis-proportionate*) meant to avoid problems that would be invited by codifying the developed case law in its actual vocabulary of ‘manifest unreasonableness.’ I.e., the

453 Widdershoven, ‘Een ervaring als staatsraad-generaal: op zoek naar een rechtsbeginsel’, 89.

454 Gerards, ‘Meer rechtsbeginselen in de Awb?’.

455 R.J.N. Schlössels, ‘Kroniek beginselen van behoorlijk bestuur 2010’, *Nederlands Tijdschrift voor Bestuursrecht* 2010, nr. 7 (2010).

456 Awb article 3:4/2. To be sure, the thesis’ use of the term ‘hardship clause’ means to label this discussion in a way that it can be easily and recognizably referred to later on. It is not the colloquial term used in scholarship.

457 de Waard, ‘Proportionality in Dutch administrative law’, 114.

construction of a legal provision that prohibits manifest unreasonableness could suggest that a bit of unreasonableness is still palatable, whereas an instruction to act reasonable would invite *more* judicial scrutiny than the Legislator envisioned.⁴⁵⁸ The idea was to continue the test as one that assesses whether there is a ‘not unreasonable’ relationship between legal *purposes* and administrative *means*, without coming too close to assessing the merits of those means itself.⁴⁵⁹ The Benefits Scandal led to fundamental discussions about the adequacy of the provision in both its roles: as instructive for proper State (administrative) behaviour and for judicial assessment of the same.

In administrative judicial traditions since then, this is indeed how the ‘proportionality’ test was applied by the Administrative courts.⁴⁶⁰ Only in cases where administrative bodies were assigned explicit discretion to balance interests, and only when the outcome of such exercises were so unreasonable that ‘no reasoning is considered to have taken place’ was a decision deemed ‘disproportionate.’⁴⁶¹ This historical tradition is being departed from in 2022.⁴⁶² Judges, and so, administrative bodies, will be expected to perform a material proportionality test that is more in line with European (EU and Council of Europe) reasoning traditions: to qualify a (planned / taken) decision’s appropriateness, necessity, and proportionality. Also when administrative bodies’ discretionary balancing space is severely limited in law.⁴⁶³

But it can be questioned how smooth that transition will be.⁴⁶⁴ What substance will administrative bodies give to their test, and where will they find the guidance to do so?⁴⁶⁵ In the years that the Benefits Scandal unfolded, neither the Judiciary nor the Tax Administration made use of the principle of proportionality. Discussions since then among judges and legal scholars reveal fundamental unclarity about whether

458 de Waard, 14–15.

459 Which would too easily amount to ‘messing with’ the actions of a democratically chosen government de Waard, ‘Proportionality in Dutch administrative law’.

460 That is not to say that administrative judicial procedures have been static: see section 4.3.1.2 for some developments that meant to make the procedure more meaningful for individuals.

461 K J De Graaf en A T Marseille, ‘Exit willekeurstoets. Bestuursrechterlijke toetsing aan het evenredigheidsbeginsel na 2-2-’22’, *Ars Aequi* 2022, nr. April : 307.

462 Raad van State, ‘Aanbeveling aan bestuursrechter: pas rechterlijke evenredigheidstoets aan (ECLI:NL:RVS:2021:1468)’, *Raad van State* (blog) (Raad van State), last consulted 6 February 2022, <https://www.raadvanstate.nl/actueel/nieuws/@126011/conclusie-evenredigheidstoets/>; De Graaf en Marseille, ‘Exit willekeurstoets. Bestuursrechterlijke toetsing aan het evenredigheidsbeginsel na 2-2-’22’.

463 De Graaf en Marseille, ‘Exit willekeurstoets. Bestuursrechterlijke toetsing aan het evenredigheidsbeginsel na 2-2-’22’.

464 To be sure, this is not to say that the transition *should* be smooth at all – it is likely to upstart the system, which can certainly be argued to be a good thing.

465 Marseille and Brenninkmeijer for example refer to Scheltema, who argued that the principle means that the larger legal system (including that of the administration) should not make ‘unreasonable’ demands of citizens, nor should it lead to ‘unreasonable’ results. Marseille en Brenninkmeijer, ‘Een dialoog met de Raad van State na de toeslagenaffaire’.

they should have, and about the principle and its codification itself. Authors on one side argued that neither the Tax Authority nor the Judiciary had discretion to apply the Awb's general (dis)proportionality test because the Childcare Benefits Act, the special primary law that the Tax Authority derived its policy making powers from, purposefully lacked such a clause to express an envisioned prohibition to use the discretion.⁴⁶⁶ Others argued that the Awb's hardship clause was never meant to be overruled by more subject specific, nor more recent, primary administrative laws. In fact, by not explicitly referring to the Awb's provision nor to a bespoke hardship valve, the Benefits Act *illegally* restricted the discretionary space of decision makers.⁴⁶⁷ Yet others, baffled by these discussions, have argued that *all* State employees and members of the Judiciary have the duty to *always* be prepared to balance an individual's interests in light of fundamental legal principles of good faith, proportionality, and subsidiarity. To quote legal scholar Pessers: "If you are not schooled—through education or institutional training—to address such questions, a job anywhere in the *trias politica* is not for you."⁴⁶⁸

Administrative law scholar Scheltema, concerned that citizens risk to suffer a similar lack of legal protection in the future, placed his bets on further codification. An addition to the general hardship clause in the Awb would allow (judges, in Scheltema's proposal) to *apply* the law in the name of justice instead of needing to argue against a law's "inner values or reasonableness." The amendment reads "[t]he application of a legal provision needs to be differentiated (sic) also to the extent that the provision cannot be applied, in cases where special circumstances that depart from normal patterns (sic) require to do so, and the law's intent was not to ignore these circumstances."⁴⁶⁹ His earlier proposal to augment the Awb's general hardship clause was somewhat different: "In cases where the application of a legal rule leads to disproportionate negative consequences in light of a rule's aims, through no fault of the interested party, the rule may be diverted from to the extent that is needed to apply it in a more balanced fashion."⁴⁷⁰ That proposal received a decidedly negative Government response. Citing recent restrictions to the discretionary space in the politically sensitive domain of immigration law, they argued that some policy domains

466 van den Berge, 'Bestuursrecht na de toeslagenaffaire'.

467 van den Berge.

468 Ten Hooven, 'De verzorgingsstaat is grimmig geworden', 16; Pessers, like Ten Berge, cites Aristotle: law is not just a set of (tough) rules but an art: 'the art of the good and the fair.' van den Berge, 'Bestuursrecht na de toeslagenaffaire'.

469 "Differentiatie in de toepassing van een wettelijke voorschrift is, ook in afwijking van dat voorschrift, geboden indien bijzondere omstandigheden in afwijking van het normale patroon dat verlangen, en de wet niet beoogt deze bijzonder omstandigheden buiten beschouwing te laten." M. Scheltema, 'Een wet van Meden en Perzen? Geen onwrikbare wet in het hedendaags bestuursrecht.' In: Wetgeving en uitvoering Nederlandse Vereniging voor Wetgeving' (Nederlandse Vereniging voor Wetgeving, 2021), 56.

470 Scheltema, 'Wetgeving in de responsieve rechtsstaat', 128.

required reduced, rather than enlarged possibilities for identifying ‘special cases.’⁴⁷¹ The amended proposal received critique from different sides, and for different reasons. Zijlstra argued the adjusted provision could invite decisional arbitrariness, and was also not necessary: in cases that pertain punitive administrative sanctions that lead to (severe) disproportionate hardship, Human Rights law’s ‘criminal charge’ protections could simply be invoked.⁴⁷² The argument however opens up to further discussions about how subjects of administrative sanctions are typically and problematically left *without* fundamental legal protections throughout their procedures in the punitive administrative law regime, as the second case illustration (section 4.2.3.6) will illustrate. The first case illustration (section 4.2.3.5) pulls into doubt that seeking a solution in enlarged discretionary policy space in special laws themselves, rather than in the Awb, is the better way to go. With reference to those cases, Bröring and Tollenaar consider how “vague laws lead to vague, unaccountable decisions.”⁴⁷³

These discussions, argues Van den Berge, are about the *character* of Dutch Administrative law. The question is whether it should be seen a legal expression of public authority, or an authoritative expression of principled justice: a more open-ended approach.⁴⁷⁴ He argues that any attempt to safeguard administrative justice in a final manner cannot be done via provisions in a law that is, all things considered, part of the system that produces the injustices in the first place. The adjudication of justice in such cases can only be “based on reasons that lie outside of it,” such as in unwritten, universal principles.⁴⁷⁵ With that, we return to the discretionary space of civil servants, and to them: the decision makers and explainers this thesis means to support with progressive, written instruction, located somewhere in law where they can come to full expression while allowing for their further development.

4.2.2 Civil servants as interactional partners: expectations and concerns (‘the who’)

Parliamentary discussions around the enactment of the Awb (mid 1990s) expressed the aspiration that administrative bodies would up their effort to make citizens’ interests count in individual decisional procedures. To not simply assume that these were already accounted for through their democratic vote. To act, communicate, and interact with citizens in responsive, rather than predetermined ways.⁴⁷⁶ Then and now,

471 Herman Bröring en Albertjan Tollenaar, ‘Menselijke maat in het bestuursrecht: afwijken van algemene regels’, in *Grensoverschrijdende rechtsbeoefening: Liber amicorum Jan Jans*, edited by K. J. de Graaf e.a. (Zutphen: Paris, 2021), 2011.

472 S.E. Zijlstra, ‘Voorwaardelijke opzet van de wetgever: enkele kanttekeningen bij het preadvies van M. Scheltema’ (Nederlandse Vereniging voor Wetgeving, 2021), <https://www.nederlandseverenigingvoorwetgeving.nl/wp-content/uploads/2021/01/Reactie-Sjoerd-Zijlstra.pdf>.

473 Bröring en Tollenaar, ‘Menselijke maat in het bestuursrecht: afwijken van algemene regels’, 211.

474 See also Tollenaar, ‘Maintaining Administrative Justice in the Dutch Regulatory Welfare State’.

475 van den Berge, ‘Bestuursrecht na de toelagenaffaire’.

476 Parlementaire Geschiedenis Awb I, p. 39-46, Available at <https://pgAwb.nl/pg-Awb-digitaal/eerste-tranche-Awb/ii-rechtsbetrekking-bestuur-burger/>.

responsive, honest, trustworthy citizen-State relations are named as constituents of the democratic state itself. In absence of such relations, the public estranges, distrust fosters, and the State's legitimacy corrodes.⁴⁷⁷

Precisely this corrosion, for precisely these reasons, is called out after the Benefits Scandal. It comes at a time when civil servants are also foregrounded as the only possible 're-humanizers' of estranging, distrust-fostering automated administrative decision practices. In their unsolicited advisory report, The Council of State acceded the apparent need for improved bureaucratic relations if these expectations are to be met in digital times.⁴⁷⁸ Their concerns broadly overlap with those voiced in 'explanation in crisis' discussions that were the subject of Chapter 2. E.g., citizens don't understand the legal rules that exist and that are applied in their cases, they cannot participate responsibly, nor check decisional outcomes. Furthermore the Council is concerned that the extent to which the complex translation (into policy) of technologically neutrally phrased laws is currently left up to the policy-making and executive branches entirely obscures the extensive norm setting that this involves. Citizens are confronted with obscure rules and parliamentary control becomes moot (since policy rules aren't subject to their control.)⁴⁷⁹ The civil servant becomes the [only] one to turn to for decision subjects... but for their interaction to be *meaningful*, their relationships will need to be improved. The Council argues this entails the further operationalization and development of the principles of motivation and due diligence, the explication of existing 'contact' rights⁴⁸⁰ and their enhancement into a *principle* of proper administration.⁴⁸¹ But in their advice they do not focus on the initial decision maker/explainer that citizens interact with. They promote the internal Administrative review procedure as the forum to make the principles especially meaningful for.⁴⁸² Meaningful in the sense that the *review* explainer knows all there is to know about the decision, and is in the position to correct it. This, they say, follows from the public value of human dignity.⁴⁸³

477 Schlössels, "Kroniek beginselen van behoorlijk bestuur 2018"; "Hoe Hoort Het Eigenlijk? Passend Contact Tussen Overheid En Burger," 8.

478 'Ongenvraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen'.

479 'Ongenvraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen', Ch 5.

480 Such rights are many and scattered across due process-like provisions such as 'the administrative body allows decision subjects to add any missing information'; 'are informed about the information they need to provide'; 'are invited to discuss an administrative body's intent to decide negatively when it is based upon information they provided' et cetera, as well as in process rights around internal review, and in Framework Laws that describe general obligations for administrative bodies.

481 'Ongenvraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen', Section 4.3; Machteld Claessens, 'Het (on)nut van een recht op toegang tot de overheid als nieuw algemeen beginsel van behoorlijk bestuur', *Nederlands Tijdschrift voor Bestuursrecht* 2021/105, nr. 4 (April 2021).

482 How this is arguably problematic in itself is discussed at length further on.

483 'Ongenvraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen', Section 4.3.

Both the cited explainee challenges and the projected capabilities of explainers that the Council describes as ‘dignified’ were argued for well in advance of the Awb codification discussions—and fed into them as we saw. Filet describes them in his 1974 empirical study. He found that Administrative rule and process complexity were obstacles to meaningful subject participation, and that the explanatory exchange is especially meaningful when explainers are in power to change a decision.⁴⁸⁴ The persistence of the cited problems therewith begs the question of how to regulate for the envisioned improvement. At the time of writing, Dutch Parliament approved a motion that requires the Government to develop the Council’s proposed right-to-principle translations into a principle of ‘meaningful government contact.’ Administrative Law scholar Claessens wrote a critical commentary. In light of the persistence of the cited problems, and of how many but scattered behavioral instructions are already present in Administrative laws and provisions, she argues that a ‘duty of care’ might be the better instrument. A duty that demands a well described result and puts the burden of proof for reaching that result on the shoulders of public decision makers.⁴⁸⁵

But the need for improvement to meet the aims of trustworthy, meaningful relations has hardly been a secret. Are existing instructions not so clear after all, or hard to meet in practice? Arguments for the latter are much explored in scholarly literature, where the ‘dual’ moral positions of public servants are investigated. Caught between expectations of impersonal, predictable, rule-based, rational, and equal treatment and expectations of principled and conscientious responsiveness to the uniqueness of individual situations.⁴⁸⁶ Among such descriptions, there is debate about to what extent these tensions are the inevitable outcome of the phenomenon of bureaucracy,⁴⁸⁷ or the avoidable negative outcome of what is an inevitably turbulent but properly governed relation in principle.⁴⁸⁸

Writing on personal title, Government information policy advisor Borst argues that this juxtaposition of the two expectations is distracting. In a re-investigation of two famous scholarly developments of the ‘impersonal’ and the ‘conscientious’ disposition, he closes (or at least, narrows) the assumed gap that exists between them. Both late 19th century German sociologist Weber, traditionally cited for his theoretization of impersonal

484 Filet, *Kortsluiting met de bureaucratie: over participatiemogelijkheden van burgers bij het openbaar bestuur*.

485 Claessens, ‘Het (on)nut van een recht op toegang tot de overheid als nieuw algemeen beginsel van behoorlijk bestuur’.

486 A.C. Widlak en R. Peeters, *De Digitale Kooi: (on)behoorlijk bestuur door informatiearchitectuur* (Boom Bestuurskunde, 2018), 39; W J Witteveen, ‘Kafka en de verbeelding van bureaucratie’, 2010, 9.

487 David Graeber, *The Utopia of Rules: On Technology, Stupidity, and the Secret Joys of Bureaucracy* (Melville House Books, 2015).

488 Schlössels, ‘Kroniek beginselen van behoorlijk bestuur 2018’; ten Bos, *Bureaucratie is een inktvis*, 95–97.

bureaucratic rule,⁴⁸⁹ and his contemporary Dutch legal scholar Scholten, cited for his descriptions of the Judiciary as moral seekers, therewith keepers, of justice acknowledge the existence of a fundamental discretionary space in rule-based individual decisions making. A space that cannot *not* exist since all human decision making entails moral choice making and therewith ‘a conscience.’ A conscience that still needs to be reasoned, which entails an engagement with the (inevitably also debatable) norms of any particular decisional domain. In Weber’s words, “principally, every act of truly bureaucratic governance is backed by a system of rationally debatable ‘considerations’ either as subsumption under norms or the outcome of balancing end and means.”⁴⁹⁰ And where all commentaries of Scholten apply his thought to the judiciary, Borst cites Scholten saying “all legal labor is Judges’ labour,” in how it all entails to make an estimation of what a Judicial opinion should conclude: what justice demands in the case.⁴⁹¹ Borst adds how this is no different for all other *trias* members. And argues that all members’ conscientious discretionary space itself needs (more) acknowledgment.⁴⁹²

With more acknowledgment of the inevitable, fundamental discretionary space of civil servants, their work instructions and environments can be investigated for how they are, or are not, conducive to using that space ‘meaningfully.’ Bureaucracy’s vastness, Administrative legal complexity and specialization, and an abundance of detailed instruction are much named as obstacles. Both literally, and in how they promote the wrong kind of disposition. The growth and specialization of Administrative Bodies meant that civil servants became responsible for increasingly detailed, ‘abstract’ tasks in ever growing, hierarchically ordered systems. The decrease of oversight, of comprehensive responsibility, and of required engagement with the rationale of higher order rules are named as drivers of rule-following inflexibility at best,⁴⁹³ of dangerously inhumane, value-free practices at the worst.⁴⁹⁴ When decision makers cease to acknowledge personal accountability, rules and procedures cease to make

489 Borst cites Caplow’s interpretation of Weber: “Indeed, the most fundamental trait of bureaucracy is a sharp distinction between the position and the man who holds it. His office is completely separated from his home, and the hours he must spend in the office are specified. His conduct as an official is not supposed to be influenced by his personal traits or affiliations or by the personal traits and affiliations of the people he deals with. The whole purpose of the bureaucratic system is to establish working relationships that are impartial, dispassionate, predictable, and uniform” Wim Borst, ‘Mag het bestuur ook wat de rechter mag? Over de verhouding tussen bestuur en rechter (naar aanleiding van de toeslagenaffaire)’, *Ars Aequi* 2022, nr. April.

490 Borst’s translation of Weber on page 664 of ‘Wirtschaft und Gesellschaft’ (1921) Borst, 389.

491 Citing Scholten in G.J. Scholten, Y. Scholten & M.H. Bregstein (red.), *Verzamelde geschriften van prof. mr. Paul Scholten*. Eerste deel, Zwolle: Uitgevers-maatschappij W.E.J. Tjeenk Willink 1949, p. 454 Borst, 385.

492 Borst, 385.

493 Quoting Schmitt, ten Bos, *Bureaucratie is een inktvis*, 94; Widlak en Peeters, *De Digitale Kooi: (on) behoorlijk bestuur door informatiearchitectuur*.

494 Hannah Arendt, ‘Some questions of Moral Philosophy’, in *Responsibility and Judgment*, Reprint edition (Schocken, 2005), 58.

sense to both them and their subjects.⁴⁹⁵ This is Kafka's territory. The prewar chronicler was especially sensitive to the idiosyncrasies of the expanding 'administrations' of his time. In his (fictional) stories he mastered descriptions of bureaucratic alienation to the point that his name became an adjective for real life administrative procedures that are 'ridiculously unreasonable, unsound, or incongruous' (Merriam Webster); 'wildly unreasonable, illogical, or inappropriate' (Oxford Online).⁴⁹⁶

Dis-encouraging civil servants to think for themselves as a tactic of (aspiring) totalitarian regimes inspired Arendt to warn for what she regarded as an *inevitable* tendency of bureaucracies: "[I]n terms of perfect bureaucracy—which in terms of ruler-ship is the rule by nobody—courtroom procedure would be superfluous (..) When Hitler said that he hoped for the day when it would be considered a disgrace in Germany to be a jurist he spoke with great consistency of his dream of a perfect bureaucracy."⁴⁹⁷ The enactment of the human rights regime after the Second World War has meant a lot for the prevention of similar 'bureaucratic horrors,' but the persistent complaints about the meaningfulness of civil servant-citizen relations express how at their level, much progress still needs to be made.⁴⁹⁸

Some argue that consecutive Dutch governments have promoted the adherence of its bureaucratic employees to strict procedural routines, restricting possibilities for responsiveness, and their mental space for reflection, or 'value rationality.'⁴⁹⁹ The promotion of automation to deal with 'bulk' decisions is an important factor, as are the not-so-smart software systems that they need to work with – more on that later.

All in all, the idea is that civil servants increasingly lack the power, both in investigative and procedural terms, to make citizens' actual situations count and to find solutions to their problems, which are mainly *made* complex through bureaucratic procedure.⁵⁰⁰ In terms of solutions, authors disagree about the beneficence of creating more discretionary space. Some argue that more horizontal and relational approaches

495 Witteveen, 'Kafka en de verbeelding van bureaucratie' *RegelMaat* 2010, 4 (2010).

496 Kafka's writings have influenced much thinking on administrative law from the time they were published (mostly after his death). Other writers known for their descriptions of bureaucratic 'purgatory' include Stanislaw Lem, Ismael Kadare, José Saramago, Italo Calvino, Jorge Luis Borges, and David Foster Wallace—all named by the late David Graeber in his scathing critique of bureaucracy David Graeber, *The Utopia of Rules: On Technology, Stupidity, and the Secret Joys of Bureaucracy*.

497 Arendt, 'Some questions of Moral Philosophy', 58.

498 See earlier sections for general remarks on the problematically individualist character of the human rights regime.

499 Widlak en Peeters, *De Digitale Kooi: (on)behoorlijk bestuur door informatiearchitectuur*, 40.

500 Widlak en Peeters, *De Digitale Kooi: (on)behoorlijk bestuur door informatiearchitectuur*; Hilke Grootelaar en Kees van den Bos, 'De Awb vanuit een procedurele rechtvaardigheids- perspectief: hulpmiddel, hinderpaal of handvat? Macht en tegenmacht in de netwerksamenleving', in *25 jaar Awb: in eenheid en verscheidenheid*, edited by A.T Marseille e.a. (Deventer: Wolters Kluwer, 2019).

are better for dealing with citizens' concrete societal needs,⁵⁰¹ others warn that blaming dehumanizing procedures on rule-following administrators disregards them as keepers of a highly complex organized legality in a system that is built to rely on such experts.⁵⁰²

Missing in between these two arguments is an argument for better instructions for making well-reasoned use of a proportionate discretionary space. What we don't need are more 'refusing public servants': a term coined for Dutch state officials who refused to marry same-sex couples because it conflicted with their personal (mostly religious, in all cases discriminatory) principles.⁵⁰³ Discretionary administrative space does not naturally blossom, as the case discussions later also illustrate. What all identifications of persistent bureaucratic challenges do have in common is that legally and 'technically' complex knowledge and decision-making regimes put pressure on the meaningfulness of explainer-explainee relationships. Especially for explainees in precarious states, the interpretative labor of trying to make sense of their situation in bureaucratic terms is disproportionate. The next section discusses how civil servants themselves are instructed to make sense of their subjects' realities.

4.2.3 Individual Administrative decision making: basic norms and instructions ('the how')

This last section concludes the first part of the chapter: the functional characterization of the 'what, who, and how' with regard to knowledge and decision making in this first of two regulated explanation domains. The section is necessarily restricted: a comprehensive description of 'how it's done' requires book-length treatments of all applicable rules and principles and the empirical reality in which these are applied. The chapter chose to focus on the basic *legal* instructions (including the *codified* principles) that govern the information positions of decision makers, and how they are expected to build their cases. Two case illustrations are added to provide depth to the necessary insight.

501 Lukas van den Berge, 'Bestuursrecht Tussen Autonomie en Verhouding: Naar een Relationeel Bestuursrecht'.

502 R.J.N. Schlössels, 'Kroniek beginselen van behoorlijk bestuur 2013', *Nederlands Tijdschrift voor Bestuursrecht* 2013, nr. 10 (2013); Schlössels, 'Kroniek beginselen van behoorlijk bestuur 2018'; See also Meuwese, 'Grip op normstelling in het datatijdperk', 157; R.M. van Male, 'Bestuursrechtspraak bij erosie van het legaliteitsbeginsel', *Nederlands Tijdschrift voor Bestuursrecht* 2019, nr. 2.

503 Conform ECHR Eweida v. United Kingdom (2013), a Dutch High Administrative court in 2016 decided these servants may be fired. (ECLI:NL:CRVB:2016:606).

4.2.3.1 *Necessary knowledge about relevant facts*

Per the Awb, the preparation of decisions obliges administrative bodies to gather “the necessary knowledge about the relevant facts and the interests that are to be balanced.”⁵⁰⁴ Necessity and relevance are governed by the aim of the decisional process, which itself follows from underlying laws and their policy interpretations. Administrative bodies need to make clear what types, and kinds, of evidence they will accept, for example in policy rules. But as the possible relevance of all information can’t possibly be predicted, no type of information can be excluded beforehand: the principle of due diligence in Dutch Administrative Law is understood to mean that “all relevant information must be available and allowed to play a role in careful preparation of the decision.”⁵⁰⁵ Administrative bodies are expected to have made sure that decision subjects were able to contribute to the evidence building in their case.⁵⁰⁶ Yet the Awb itself is inconclusive about the distribution of burdens with regard to this task.⁵⁰⁷ And in light of how it can be highly challenging for decision subjects to participate meaningfully in the stage of fact collection, such ‘unclear’ codifications are of importance. Acknowledged challenges especially pertain to information that is collected about subjects by administrative bodies; more specifically to a lack of meaningful influence of citizens on the quality and accuracy of such information.

Especially (but not only) in social/financial support cases, much information already resides in administrative systems. But it is likely distributed over different silos, governed by different administrative bodies. These different organizational units each have their own legal grounds for fact collection, they use different methods, and create and maintain specific information policies. Some collections are maintained real-time, some are updated at discrete moments.⁵⁰⁸ All in all, the Dutch system of administrative information governance is acknowledged to be complex and ‘messy,’ which also proves burdensome for the civil servants that need to make responsible use of it.⁵⁰⁹ For most administrative decisions, information needs to be created from an

504 Awb Article 3:2

505 Schlössels en Zijlstra, *Bestuursrecht in de sociale rechtsstaat 1*, section 8.3.6, para 42–43.

506 Marion Beckers et al, red., *Motiveren, Over het motiveren van rechterlijke uitspraken*, Prinsengrachtreeks (Ars Aequi Libri, 2017).

507 Y.E. Schuurmans, ‘Bewijslastverdeling in het bestuursrecht: zorgvuldigheid en bewijsvoering beschikkingen’ (Vrije Universiteit van Amsterdam, 2006), 53; Janny Kranenburg, ‘The facts: Administrative versus Civil Law Courts,’ in *Motiveren, Over het motiveren van rechterlijke uitspraken*, edited by Marion Beckers et al, Prinsengrachtreeks (Ars Aequi Libri, 2017), 44–46.

508 Widlak en Peeters, *De Digitale Kooi: (on)behoorlijk bestuur door informatiearchitectuur*.

509 The ungoverned proliferation of ICT systems made this decidedly worse over time. As the Scientific Council warned in 2011, ‘Meanwhile, both the relevant government official and the citizens in question are unaware of the deterioration (...) Administrative reality and “real reality” can diverge quite dramatically in iGovernment, and errors can be disseminated much more quickly, making them more difficult to rectify later on. Such errors can have huge repercussions for the daily lives of individual citizens ...’ Corien Prins et al, ‘iGovernment - Synthesis of WRR Report 86 (English Version)’ (The Netherlands Scientific Council for Government Policy, 15 March 2011).

accumulation of different ‘bits’ that are therewith re-interpreted, and may themselves consist of earlier combined calculations and information. Legal and policy efforts to ‘clean up’ are continuously underway, but prove tricky. An illustration: around 2009, the ‘single authority’ information establishment system was introduced. It was meant to reduce the problem of diverging registrations of what should be ‘one truth,’ such as a subjects’ place of residence. For each type of registration, one administrative body was assigned the role of trusted establishing party. Some administrative bodies were put in charge of a combination of facts, such as a vehicle’s registration but also its insurance, or several different income aspects. The system came with the strict obligation for all administrative bodies to use the establishing body’s registrations, even if it was contested by decision subjects, and even if the information was in conflict with more recent information held elsewhere. ‘Client’ bodies in such cases were obliged to present their doubts to the establishing body, who would be in charge of (decisions with regard to) any corrective labour. Various complications of the system were foreseen by scholars at the time, and indeed established.⁵¹⁰ E.g., citizens have a hard time tracking, and maintaining, their information with all establishing administrative bodies who need to provide input for a decision.⁵¹¹ And as the Kafka Brigade investigated, some client bodies worked with a downloaded version of an establishing body’s register. This relieved them of the time-consuming burdens of querying different systems real-time.⁵¹² Fault chain reactions were the result,⁵¹³ but the reason for such faults remained a mystery as decision subjects had no way of knowing how this happened.

Decision subjects are also asked, and *in casu* obliged, to bring information to the table themselves,⁵¹⁴ including information that resides in systems that administrative bodies don’t (yet) have access to: medical files, employment contracts, bank account statements. The digitization of *these* systems makes such collection burdensome for the millions of Dutch citizens who lack digital literacy or don’t have access to electronic means.⁵¹⁵ For those whose required information resides in foreign systems, problems tend to multiply. E.g., refugees notoriously lack the ability to meet the

510 G. Overkleef-Verburg, ‘Basisregistraties en rechtsbescherming. Over de dualisering van de bestuursrechtelijke rechtsbetrekking’, *NTB* 2009, nr. 10, last consulted 27 May 2019.

511 Such input can also be ‘decisions’ in themselves, such as calculations that made use of different sources. ‘Chain decision making’ became the norm, as was also explained earlier. Overkleef-Verburg.

512 Kafka Brigade, personal conversation with Arjan Widlak

513 See also van Eck, ‘Geautomatiseerde ketenbesluiten & rechtsbescherming: Een onderzoek naar de praktijk van geautomatiseerde ketenbesluiten over een financieel belang in relatie tot rechtsbescherming.’; and Corien Prins et al, ‘IGovernment (English Version)’ (The Netherlands Scientific Council for Government Policy, 15 March 2011).

514 For example Article 4.2/2.2. documents “as required for a decision on the application as it is reasonable to expect him to be able to obtain.” and Article 4:3, stating how applicants may refuse to supply information and documents but only “in so far as their importance to the decision of the administrative authority is outweighed by the importance of protecting privacy, including the results of medical and psychological examinations, or by the importance of protecting business and manufacturing data” and not if those kinds of information are required by the underlying law.

515 ‘Ongevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen’.

Policy demands for documentation.⁵¹⁶ In all cases, information provision demands can become disproportionate.

To return to the distribution of burdens with regard to making sure the right information is available for decision making, it is important to mention that Dutch Administrative Law does not know a bespoke regime of truth finding and evidence (unlike, for example Criminal Law.) There are no clear rules with regard to what falls under the ‘information duties’ of citizens, privacy and data protection laws also apply, and all in all it can be unclear for citizens if they are obliged to comply with a request for information.⁵¹⁷ In addition, administrative bodies have considerable legal ‘discretion’ with regard to experimenting with (new) methods of fact collection, such as web scraping methods.⁵¹⁸ Before digital times, as well, fact-finding missions of administrative bodies landed in court, which then (could) lead to the establishment of jurisprudential norms.⁵¹⁹ In 2020, a seminal case struck down an entire law that enacted an automated fraud-detection regime in the social domain. The case was much discussed internationally as the first such case where fundamental rights prevailed fundamentally in ‘black box’ times.⁵²⁰ A new law that allows administrative bodies to exchange practically unlimited information with private entities is being debated in the Senate at the time of writing—and is sure to end up in court if it indeed becomes law.

516 Ranging from identity papers to evidence for being prosecuted, leading to much litigation of evidential burdens. Schuurmans, ‘Bewijslastverdeling in het bestuursrecht’, for example 77-78.

517 For example, in the first period after the decentralisation of welfare and care domains, the DPA reported on gross non compliance with privacy and data protection laws after case workers were asked to ‘use their trusted relations with citizens’ in gathering the necessary facts. ‘Verwerking van persoonsgegevens in het sociaal domein: De rol van toestemming’ (Autoriteit Persoonsgegevens, April 2016).

518 Oerlemans en Schuurmans, ‘Internetonderzoek door bestuursorganen’.

519 A seminal 1987 court case that is part of standard Dutch law students’ curriculum condemned the self-initiated spying and reporting on a woman on single household welfare by her neighbour, who worked for the municipality. ECLI:NL:HR:1987:AG5500, voorheen LJN AG5500, AC0705, AJ3785, AM9322, Hoge Raad, 12.717, No. ECLI:NL:HR:1987:AG5500 (HR 9 January 1987).

520 NJCM and others v. The Netherlands (English) ECLI:NL:RBDHA:2020:1878 (The Hague District Court 6 March 2020); The case attracted attention from the UN, resulting in an Amicus Brief by Philip Alston, UN Special Rapporteur on extreme poverty and human Rights Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights in the case of NJCM / De Staat der Nederlanden (C/09/550982/HA ZA 18/388), ‘Brief by the United Nations Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM C.s./De Staat der Nederlanden (SyRI) before the District Court of The Hague (case number: C/09/550982/HA ZA 18/388)’, 26 September 2019; For an English summary see ‘SyRI legislation in breach of European Convention on Human Rights’, last consulted 23 September 2020, <https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx>; and ‘The SyRI Victory: Holding Profiling Practices to Account’, last consulted 25 September 2020, <https://digitalfreedomfund.org/the-syri-victory-holding-government-profiling-to-account/7/>.

A last thing to mention in this section is the concept of ‘legal fictions.’ Legal fictions are used to mediate legal descriptions of social situations and the inevitably more multiform lives of decision subjects. They restrict the kinds of facts that administrative bodies need to consider. Put differently, legal fictions are created to categorize the multiform ‘truths’ of decision subjects in terms of legal relevance. How it works: a particular law or policy determines that one (set of) fact(s), sometimes with the added requirement for how these facts are established, symbolize(s) a legally relevant situation. Other information or facts are deemed irrelevant to it: they can’t be made to count.⁵²¹ E.g., a welfare law or policy may determine that two persons who spend a certain amount of days per month in the same apartment are assumed to share the financial burdens of the household, which means they are not eligible for single-household benefits—and administrative bodies get to determine the evidence they allow themselves to go on. Fighting unfair decisions in such cases can take many years, efforts and affordances.⁵²² These kinds of fictions tend to work out especially problematically for persons and groups in precarious states. E.g., sharing a friend’s roof is a perfectly humane coping strategy in case of temporary homelessness. But a legal fiction such as the one in the example means only those who can afford to live outside welfare scrutiny can afford to do so. A weaker version of a legal fiction is a legal ‘assumption.’ These can be disproven by decision subjects. In line with what was said about how decisions under one policy are also relevant for decisions of other administrative bodies, this difference matters a lot, as will the conditions with which an assumption can be negated. The second case illustration in section 4.2.3.6 testifies to the very complex situations the interplay of legal fictions and assumptions (in combination with various other Administrative particularities) can amount to. The cases were referred to by the Council in their argument against the State’s broad exception clause for the GDPR’s article 22, see section 4.2.3.3.

4.2.3.2 Necessary knowledge about interests to balance: pro’s and cons of the ‘specialty’ principle

The second part of Awb’s article 2:3 pertains the obligation to gather the necessary knowledge about interests that need to be balanced. The obligation needs to be read in combination with the first paragraphs of the Awb’s hardship clause (3:4, under one): administrative bodies are obliged to weigh those interests that are ‘directly related’ to the decision, within the bounds of legal rules or the discretionary nature at hand. This

⁵²¹ Schuurmans, ‘Bewijslastverdeling in het bestuursrecht’, 86, 174–76.

⁵²² For example, a (black) Pastor on single household welfare was robbed of his income after case workers, suspicious after finding female clothing in his apartment and not believing his account that the clothing belonged to a Belgian family member who stayed with him when she visited, asked him what the neighbours would say when the Municipality would ask them about their relationship. In all honesty, the Pastor answered that they would probably assume that she was a love interest. The supreme Administrative court eventually found that the Pastor was wrongly burdened with the burden of evidence as well as the risk that the imputation this amounted to. ECLI:NL:CRVB:2014:1035, Centrale Raad van Beroep, 13-4228 WWB, No. ECLI:NL:CRVB:2014:1035 (CRvB 20 March 2014).

should be seen in light of the prohibition to use decisional powers for other aims than they were attributed for (*détournement de pouvoir*.) The regime is also referred to as the codification of the ‘specialty principle,’ which refers to distribution of specialized decisional powers over different administrative bodies.

What interests are relevant to an administrative body’s specific legal authority may be explicit in underlying laws, but they are also be *derived* from it in light of a law’s aims and objectives. And although ‘alien’ interests don’t have a place in such reasoning *in principle*, an administrative body can still take them into account if they can ‘prove’ their relevance with good reasons.⁵²³ It therefore matters how decision makers are instructed to make use of the law, and of the written and unwritten principles that also apply. The interplay of the principles of due diligence and motivation potentially allow a much broader discretionary space. Without clear obligations to make use of this space, the regime easily limits the balancing that may *need* to be done when the interplay of different administrative bodies’ decisional powers leads to disproportionate hardship for subjects. An important consideration in light of the very scattered decisional landscape.⁵²⁴ Decisions of one domain become part of decisions in another, but the negative effects that a decision made under authority A will have on a decision subjects’ position under authority B, are ‘irrelevant’ interests. Less emphasized in literature is how the regime in place for the prevention of power abuse especially restricts the understanding that can be gained about *less clear* abuses of power: those that only become visible when intersectional, marginalizing dynamics are investigated and made to count. The Benefits Scandal generated discussion on both these points. As was cited in Chapter 2, Administrative judges deplored the fact that they had not taken an interest in the effects of their decisions on the other Administrative decisional processes the victims were subject to, keeping disproportionate hardship out of view.

4.2.3.3 The Dutch Article 22 exception clause: a codified loss of functionality?

The GDPR is directly applicable in Member States. Still many States enacted ‘implementation laws’ to guide local actors in their implementation of the complex new regime. Although slight differences in approaches of Member States tend to creep in through such laws, The Netherlands was the only country to undermine a constitutive aspect of the GDPR’s article 22, which governs the prohibition of ADM.⁵²⁵ The GDPR requires that for each application of ADM a State wants to make possible, a specific law with bespoke safeguards needs to be enacted. Instead, The Netherlands legislated a broad exception clause on the basis of ‘public interest’ in the GDPR implementation

523 “How does one balance interests?” Etjo Schrage in Marion Beckers et al., eds., *Motiveren, Over het motiveren van rechterlijke uitspraken*, Prinsengrachtreeks (Ars Aequi Libri, 2017), 29.

524 Schlössels and Zijlstra, *Bestuursrecht in de Sociale Rechtsstaat 1* para 8.3.2, under 15-16.

525 Gianclaudio Malgieri, ‘Automated Decision-Making in the EU Member States Laws: The Right to Explanation and Other “Suitable Safeguards”’, *Computer Law & Security Review* 35, nr. 5 (forthcoming , Available at SSRN: <http://dx.doi.org/.2139/ssrn.3233611> 2019): 8.

law. The clause is available to all administrative bodies, to serve all of their aims, as long as they don't engage in 'profiling'.⁵²⁶

The Government defended the exception by arguing that not all ADM involves 'using group characteristics against an individual'—a flawed understanding of the wrongs and harms that any 'profiling' can enable—or 'processing personal data with a risk of ensuing discrimination' and that human intervention is of no added value in such cases.⁵²⁷ The Council of State, referring to a notorious kind of simple yet harmful ADM (featured in section 4.2.3.6) rebutted that the Government's 'simple' cases have led to disproportionate and unreasonable consequences for many citizens.⁵²⁸ The clause stood: the Government argued that automation in such cases is not necessarily to blame: the decision makers may have been bound by law and [their own] policy rules.⁵²⁹

The response ignores how the decision makers in the cited cases expressed a convoluted legal interpretation and application, and how the automation was a crucial factor for harm delivery as described in the reports cited by the Council.⁵³⁰ It also ignores the additional restrictions on interest balancing that the clause promotes. This hardly needs to be explained again in this chapter, but it is useful to quote Dutch Professor De Mulder. In 1993, he voiced an "educated guess" that it will be hard to program the human-intuitive legal relevance of any given situation, which could establish a "loss of functionality" through automation.⁵³¹ Since then, evidence of how automation is functionally deployed to effectuate generally harsh, and effectively discriminatory socio-economic support policy in the Dutch 'post welfare' state has piled up, leading the UN special rapporteur on extreme poverty to issue an *Amicus Brief* to the Dutch court that treated the earlier named fraud-detection law case.⁵³²

For the purposes of the thesis, the legislated exception testifies to a risky combination between the Governmental push for automation in the Administrative domain and their inclination to avoid regulating for well-known risks. This matters for thinking about the

526 'Uitvoeringswet Algemene verordening gegevensbescherming' (2018), Article 40, under 1, https://www.eerstekamer.nl/behandeling/20180522/publicatie_wet/document3/f=/vkoj2ezcplyz.pdf.

527 'Wetsadvies W03.17.0166/II - Uitvoeringswet Algemene verordening gegevensbescherming' (Raad van State, 2017), 21.

528 'Wetsadvies W03.17.0166/II - Uitvoeringswet Algemene verordening gegevensbescherming', 21.

529 'Wetsadvies W03.17.0166/II - Uitvoeringswet Algemene verordening gegevensbescherming', 42.

530 'Gegijzeld door het Systeem. Onderzoek Nationale ombudsman over het gijzelen van mensen die boetes wel willen, maar niet kunnen betalen' (Nationale Ombudsman, 2015); 'Weten is nog geen doen. Een realistisch perspectief op redzaamheid' (Wetenschappelijke Raad voor het Regeringsbeleid, 2017).

531 R.V. De Mulder, Beschikken en Automatiseren: Preadviezen voor de Vereniging voor Administratief Recht, *Nederlands Juristenblad* 1993, no. 17.

532 Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights in the case of NJCM /De Staat der Nederlanden (C/09/550982/HA ZA 18/388), 'Brief by the United Nations Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM C.s./ De Staat der Nederlanden (SyRI) before the District Court of The Hague (case number: C/09/550982/ HA ZA 18/388)', 26 September 2019.

necessary further development, explication, and codification of the principles of due process and motivation. The next section introduces the type of reasoning and balancing that human decision makers are typically assumed to do in service of the principles.

4.2.3.4 *Imagining categories of people: some words on legal reasoning*

Written explanations (justifications, reasons, motivations) can be typified as a ‘linguistic act’.⁵³³ What happens (or should happen) in language-based explanation is studied in many disciplines: argumentation theory, communication theory, epistemology, logic, mathematics; but also psychology and sociology.⁵³⁴ All these literatures can help to understand what happens between people who use language to communicate, and what makes better or worse practices in terms of interpersonal understanding.⁵³⁵ Methods, data, and assumptions from these fields are engaged with by scholars from different disciplines on the subject of ‘legal reasoning.’⁵³⁶ Researchers in pursuit of clues about required explainability of ADM in turn source works from all the mentioned fields, adding their take on what legal reasoning is and—at minimum—needs to be, and reaching very different conclusions as was discussed in Chapter 2. This section does not provide an overview of all such notions and applications but offers a basic understanding of what is broadly understood to happen ‘linguistically’ in law-based, law-abiding, reasoning and (therewith) adds a clue about quality of information positions that ‘legal reasoners’ need to have to engage in the practice. The brief exposé helps to understand the ambition of Administrative explanation rules that are treated in section 4.3.

Theoretizations of the phenomenon of ‘legal reasoning’ are typically focused on judicial reasoning rather than initial justifications of Administrative decisions. But the reasons that administrative bodies *do* deliver in first instance are still expected to engage with the same notions, and keep to the same standards.⁵³⁷ Ceteris and Kloosterhuis distinguish three traditions: logical, rhetorical and dialogical approaches. The logical approach (itself consisting of different types) is typically described

533 Frederick Schauer, “Giving Reasons,” *Stanford Law Review* 47, no. 4 (April 1995): 634; Mireille Hildebrandt, “Law As an Affordance: The Devil Is in the Vanishing Point(s),” *Critical Analysis of Law* 4, no. 1 (2017): 119.

534 Much cited for bringing some of these insights to bear for XAI is Miller; Tim Miller, ‘Explanation in Artificial Intelligence: Insights from the Social Sciences’, 22 June 2017, <https://arxiv.org/abs/1706.07269>.

535 The linguistic approach is certainly not the only one: other disciplines study of explanation too. E.g., the fields of visual communication has pertinence in a world where distanced, digital, distributed explanations have become a necessity. Non-verbal communication is a field of study, and fiction deals with ‘explanation’ too – Kafka is only one well-known example.

536 E.g. legal theory & philosophy, sociology, artificial intelligence .. E. Feteris and H. Kloosterhuis, “The Analysis and Evaluation of Legal Argumentation: Approaches from Legal Theory and Argumentation Theory,” *Studies in Logic, Grammar and Rhetoric* 16 (2009), <https://dare.uva.nl/search?identifier=c5ea4a59-da01-43b3-88e5-02e7169502a7>.

537 Although (as we will see) this is rather implicit in Administrative explanation rules.

as requiring a ‘generalizable rule’ as a baseline.⁵³⁸ Critical of understanding legal reasoning as logical is the earlier cited Cohen,⁵³⁹ and also Hage, Leenes, and Lodder in their investigation of legal knowledge systems, arguing to take procedural law into account as well as the substantive rules when understanding the quality of legal reasoning: “Legal conclusions are not true or false independent of the reasoning process that ended in these conclusions. In critical cases this reasoning process consists of an adversarial procedure in which several parties are involved. The course of the argument determines whether the conclusion is true or false.”⁵⁴⁰

Still it helps the brief and simplified presentation of legal reasoning in this paragraph to regard a legal decision as the application of a general rule to a specific situation. The explanation of how a general rule *should* apply to a specific situation is typically described as the creation of a new ‘generalizable rule.’ That is, in service of the principle of ‘equal treatment,’ like situations need to be treated in like ways, and so there is a need to typify any situation at hand as one or another.⁵⁴¹ In Schauer’s well-known development: the justification of an individual application of a rule requires the ‘imagination’ of an individual’s situation as a new, generalizable category: a category that fits the larger rule and the ideas behind it *in principle*, and that can be used again should a like situation present itself. He continues to say that the legal demand for giving reasons is therewith inevitably in tension with doing a person’s individual situation ultimate justice.⁵⁴² In the more alarming words of a current Dutch Minister of State: “any kind of standardization threatens to pervert the principle of ‘equal treatment of equal cases’ to the need to find cases that can be treated the same.”⁵⁴³ However, the demand to use or imagine a category, as a specific kind of normative exercise also prevents that no negotiated or negotiable norm is used at all and arbitrariness rules. The necessary engagement with the underlying law itself, the reasoning about what a law was meant to do makes that law itself insightful in terms of the justness it does or does not promote and allow for. Law’s progress depends on this. And explanation rules are of influence on the type of engagement decision makers will seek to do.

Where the translation of a general rule to an individual situation (and vice versa) is one thing that typically happens in legal reasoning, the balancing of interests, importantly, is another. The idea is that interests of (and between) decision subjects, the state, private entities are ‘weighed’ to find out which one deserves to trump the other. Some

538 E. Feteris and H. Kloosterhuis, ‘The Analysis and Evaluation of Legal Argumentation: Approaches from Legal Theory and Argumentation Theory’, *Studies in Logic, Grammar and Rhetoric* 16 (2009): 312.

539 Cohen, ‘The Ethical Basis of Legal Criticism’.

540 Jaap C. Hage, R. E. Leenes, and Arno R. Lodder, ‘Hard Cases: A Procedural Approach’, *Artificial Intelligence and Law* 2, nr. 2 (1994): 113–67.

541 Schauer, ‘Giving Reasons’, 635, 641–42.

542 Reason giving as “the kin of abstraction, of rule-based decision making, and of decontextualization.” Schauer, 658.

543 Minister of State Herman Tjeenk Willink, cited in Widlak en Peeters, *De Digitale Kooi: (on) behoorlijk bestuur door informatiearchitectuur*.

argue that the methods used in balancing are not so different from general-to-specific reasoning that was described above.⁵⁴⁴ Both activities entail the qualifying of facts and interests *in light of* a relevant rule. And the balancing itself necessitates something like the imagining of generalizable categories: the quest is for a ‘generalizable way’ to make distinctions between interests, saliently between principally *equal* interests,⁵⁴⁵ which entails the ‘imagining’ of situations in which one interest would trump the other. The linguistic acrobacy that can go into such reasoning is criticized for how it ceases to make sense. E.g., interests are given different weight under the guise of respecting their equality.⁵⁴⁶

What follows is a case illustration of how the above-described legal reasoning is not naturally engaged with by administrative bodies in their use of discretionary space, even if they are asked to do precisely that.

4.2.3.5 Case illustration 1: ‘tailor made’ decision making in absence of reasoned categories

Two case illustrations close this part of the chapter. They illustrate a range of aspects that were discussed to make them more insightful. The first case treats municipal use of broad discretionary space for policy making and how this resulted in (among other things) a problematic lack of justification of State action. The second illustrates how very complex (and unsafe) administrative landscapes established a consequence of especially one particular administrative body’s choices with regard to their single information registration authority.

In 2015, the policy domains of juvenile care and social security (welfare, unemployment and disabilities) were brought under one legal regime. At the same time, the execution of the new regime was decentralized.⁵⁴⁷ This major overhaul was introduced above in section 4.2.1.2, as an example of administrative ‘aliveness’ to wrongful political climates. That section functions as a backdrop to this one, that zooms in on a particular problematic consequence of the primary laws in which the regimes were laid down, which is the creation of obscure policies. The laws prescribed very general rules and aims per domain, and attributed a large discretionary policy making space to municipal levels. Municipal administrative bodies were expected to become expert navigators of the legally integrated regimes. They were asked to make

544 H. Kloosterhuis and Carel Smith, “Hoe Werkt Het Juridisch Syllogisme?,” *Ars Aequi* 2019 (February): 155–59.

545 Such as those of fundamental rights: freedom of expression v. privacy, etc..

546 Bart van der Sloot, ‘The Practical and Theoretical Problems with “Balancing”: Delfi, Coty and the Redundancy of the Human Rights Framework’, *Maastricht Journal of European and Comparative Law* 23, nr. 3 (June 2016): 439–59.

547 “The biggest administrative overhaul since the Second World War.” Tweede Kamer, vergaderjaar 2016-2017, 34 477, nr. 10 p. 15, cited by ‘Terug aan tafel, samen de klacht oplossen: Behandeling klachten over zorg, jeugdhulp en begeleiding naar werk’.

use of their local expertise to come to responsive, tailor-made policies and decisional processes. To find solutions to citizens' actual problems rather than perform eligibility checks of citizens' support applications. Not unimportantly, the expectation was also that municipalities would make the most of bargaining cost-reducing contracts with local, private service providers.⁵⁴⁸

One of the laws that was amended to fit these purposes was the 2007 Social Support Act (which became the 'WMO 2015', WMO hereafter). The Act governs eligibility for material support such as a wheelchair, domestic help, or transportation. The Explanatory Memorandum argued that the envisioned success of the 'tailor made' regime would depend on proper municipal investigations of the help request and of the legal possibilities to meet the request (indeed necessitating skillful navigation of the integrated regime.)⁵⁴⁹ It also called on municipalities to improve the standard of motivation of decisions, which had been found lacking in judicial procedures when the decisions were made on national level.⁵⁵⁰

To effectuate the envisioned integral approach, municipalities created multidisciplinary 'neighborhood task forces' in their policies: teams with expertise on different aspects that a typical case comprised of (financial, social, medical, local).⁵⁵¹ Upon a request for social support, a task force would be deployed to chart the individual's needs, capabilities, and means. But in absence of national guidance, adequate budget, and robust monitoring, a very complex landscape of responsibilities and legal obligations established: between different municipal levels, between them and delegated private entities, and between both parties and citizens. It quickly became hard to navigate itself.⁵⁵²

The first comprehensive 'post-decentralization' report of the National Ombudsman, based on two years of complaints procedures recorded a range of fundamental issues,⁵⁵³ and much research followed that brought problematic practices to light. Especially relevant in light of the thesis's subject is the lack of reasoned use of the broad discretionary space. For example, citizen's initial support requests were not treated formally or logged in their case files, but recorded only as 'notifications.' These were

548 'Terug aan tafel, samen de klacht oplossen: Behandeling klachten over zorg, jeugdhulp en begeleiding naar werk'; Tollenaar, 'Maintaining Administrative Justice in the Dutch Regulatory Welfare State', 6–8.

549 Tweede Kamer der Staten-Generaal, 'Memorie van toelichting Regels inzake de gemeentelijke ondersteuning op het gebied van zelfredzaamheid, participatie, beschermd wonen en opvang (Wet maatschappelijke ondersteuning 2015) (kst-33841-3)', 13 January 2014, <https://zoek.officielebekendmakingen.nl/kst-33841-3>.

550 Staten-Generaal, 10.

551 Lukas van den Berge, 'Bestuursrecht Tussen Autonomie en Verhouding: Naar een Relationeel Bestuursrecht', 220.

552 Tollenaar, 'Maintaining Administrative Justice in the Dutch Regulatory Welfare State'.

553 The report recounts of decision subjects facing issues of complexity, inscrutability, and social pressure. 'Terug aan tafel, samen de klacht oplossen: Behandeling klachten over zorg, jeugdhulp en begeleiding naar werk'.

then investigated by neighborhood task forces in an informal setting whose very formal status and consequences were frequently unknown to applicants: effectively these were secret policy rules.⁵⁵⁴ E.g., when citizens asked for remuneration of informal care that was provided to them, task forces investigated the informal care burden of the applicants' social contacts to check whether these were 'above common standards' without explicating what these standards are.⁵⁵⁵ The result of such initial task force investigations was a yes/no admittance to the *actual* application procedure. The 'pre-decision' would however not be issued in the form of a Awb decision that falls under the justification and appeals regime, and so a negative answer would simply block the road to support.⁵⁵⁶ Tollenaar discusses the 'notification' regime as a harm to citizens access to justice and a consequence of legislative choice to informalize and formalize at the same time: the notification as start of application procedures was put to law whereas the form and procedure for dealing with notification was not.⁵⁵⁷ Proper eligibility decisions themselves were frequently phrased in unprecise terminology, which in itself prevented proper legal follow-up. 'Clean house' decisions are a notorious example of this. Applicants would be told that their eligibility for a 'clean house' was established, which phrasing gives no reason to seek appeal. But the decision would not include an estimation of how many cleaning hours the result was expected to take. That calculation was left up to private contractors, but since these are not administrative bodies themselves *their* decisions are excluded from Administrative appeal.⁵⁵⁸

With reference to the chapter's earlier remarks on the distribution of powers and the influence of political climates, the resistance to correction that municipal administrative bodies have demonstrated is interesting to note. Municipalities were repeatedly condemned for their behavior in court procedures. They would be obliged to perform an investigation anew, and to describe an applicants eligibility precisely in decisions.⁵⁵⁹ But in 2019, 'clean house' decisions were still made. Some municipalities admitted that the budget-relief that such procedures afforded them (and the minimalist cleaning hours that resulted from them) were well worth the cost of incidental court

554 'Terug aan tafel, samen de klacht oplossen: Behandeling klachten over zorg, jeugdhulp en begeleiding naar werk', 29 and throughout.

555 Informal carers were for example asked whether they would retract their care if they would not be compensated. Marjolijn De Boer en Sylvana Van den Braak, 'Verhuizen om wél hulp te krijgen?', *De Groene Amsterdammer*, 25 September 2019.

556 A.T. Marseille en M.F. Vermaat, 'Burgers op zoek naar rechtsbescherming in het sociaal domein', *Handicap & Recht* 1, nr. 1 (June 2017): 11.

557 Tollenaar, 'Maintaining Administrative Justice in the Dutch Regulatory Welfare State', 9–10.

558 Marseille en Vermaat, 'Burgers op zoek naar rechtsbescherming in het sociaal domein', 14; Bröring en Tollenaar, 'Menselijke maat in het bestuursrecht: afwijken van algemene regels', 11–12.

559 Yolanda De Koster, 'Wisselend succes na mediation sociaal domein', *Binnenlands Bestuur* (blog), 16 November 2018, <https://www.binnenlandsbestuur.nl/sociaal/nieuws/wisselend-succes-na-mediation-sociaal-domein.9601304.lynx> See for example the administrative supreme court for social support matters: ECLI:NL:CRVB:2017:3633.

proceedings.⁵⁶⁰ The Minister of Health at the time proposed to create a rule that would allow the decisions as long as a clean house was what applicants would indeed end up with. The proposal received critique for how it ignored and undermined the judiciary's conclusion that such decisions amount to unacceptable legal uncertainty.⁵⁶¹

4.2.3.6 *Case illustration 2: administrative truths & real-life effects of phantom vehicle license registrations*

The second case illustration was used in the Council of State's argument against the broad legal exception for administrative ADM in the GDPR's implementation law.⁵⁶² The summary below discusses the most salient aspects of relevance to the chapter. The discussion is largely based on previous research by the author.⁵⁶³ This included several interviews with representatives of the Netherlands Vehicle Authority⁵⁶⁴ (RDW); studies of the paradigm's legal history and of statements of reasons from all related administrative bodies; interviews with victims and insurance administrative bodies; FOI request results; court case attendance notes of all judicial instances, and ECtHR case law.

Vehicle license registrations: in law

In the Road Traffic Act ('WVW') article 1, under 3, a person whose name is registered to a vehicle in the Vehicle License Register (Register and registree from hereon) is assumed to be either the vehicle's owner, or its 'guardian.'⁵⁶⁵ When no actual perpetrator of a crime or misdemeanor that involves a vehicle can be established, the registree is the starting point of the investigation. In principle, Criminal Law protections apply: the law's *assumption* of ownership protects the *presumption* of innocence until proven guilty. The paradigm afforded protection to large numbers of

560 B.J. van Ettehoven en A.T Marseille, 'Afscheid van de klassieke procedure in het bestuursrecht?', in *Afscheid van de klassieke procedure?*, 2017de-1ste dr., vol. 147, Handelingen Nederlandse Juristen-Vereniging (Wolters Kluwer, 2017).

561 Vermaat in De Boer en Van den Braak, 'Verhuizen om wèl hulp te krijgen?'

562 'Wetsadvies W03.17.0166/II - Uitvoeringswet Algemene verordening gegevensbescherming', 21, and footnote 101.

563 Two phases of research were engaged with by me in the course of my evening studies for bachelor's and master's degrees at the University of Amsterdam, (UvA) while I was employed at the legal aid firm Struycken Advocaten. The products reside in the UvA's archival systems. They are not publicly accessible, but can be retrieved on special request or (more easily) sent upon request by the author.

564 'Netherlands Vehicle Authority' is the English translation that this administrative body uses for itself on its English website. The thesis used their own words, although different translations pop up in English references, e.g. Road Traffic Agency. <https://www.rdw.nl/en/>.

565 This looks, and is, complicated in itself: Dutch law knows many possible relations of persons to vehicles. When driving a vehicle, one is legally the 'posessor' and also the 'assumed owner,' a claim that can be negated. A registree of a vehicle needs to be neither the owner nor the possessor, but they are the starting point for any investigations with regard to a specific vehicle, and some responsibilities are specifically attached to their status as registree – this will be discussed in the case illustration.

drug addicts and otherwise vulnerable persons who were bribed, coerced or tricked by criminals into having vehicles registered in their name so that the vehicles could not be traced to who actually used them.⁵⁶⁶

In a set of later laws governed by the WVV, registrees are tasked with ‘vehicle obligations:’ insurance, road-tax, and periodic technical maintenance (‘APK’). That did not disculpate actual owners or drivers of vehicles for these obligations, but helped to ensure that the obligations were met in the first place. ‘Helped’ to ensure, since registrees are still only assumed to be the owner/guardian, and could disculpate themselves.

In the still younger Administrative Enforcement of Traffic Misdemeanors Act (‘WAHV’), legal enforcement for “simple transgressions, that are easily proven”⁵⁶⁷ were transferred from the Criminal to the Administrative legal system. The law also allowed for automated establishment through e.g. traffic camera’s. Owners and actual perpetrators for WAHV transgressions are no longer pursued: the registor is addressed.⁵⁶⁸ Enforcement for the vehicle obligations APK and insurance were added to the regime later on. Compliance with these is established through cross-register checks that were later automated.

The WVV’s definition of ‘registor’ was acknowledged in WAHV’s article 8, under c. A registor who can prove they were not the owner or guardian is absolved. For this, registrees need to provide evidence of unauthorized use (joyriding), or an exonerating document that proves the vehicle registration was struck before the incident or cross-register check took place. The document needs to come from the Netherlands Vehicle Authority (Rijksdienst Wegverkeer: ‘RDW’), who is singularly authorized with regard to maintaining the Register.

The RDW’s policy rules about the Register are laid down in the Vehicle License Regulation (‘Kentekenreglement’). When a registor wants to be delisted, they need to request a mutation and submit a limited list of prescribed documents. The vehicle will then be registered to another person, or a company, or is struck altogether in case of e.g. total loss or export. If a person cannot produce a particular document, the RDW is still

566 In a 2009 Supreme court case (ECLI:NL:HR:BI7044 / HR 29-09-2009, nr. 07/12516) the strong status of ‘assumption’ was confirmed, and the appellate court’s judgment that the registor should be treated as the owner because they had themselves to blame for the assumption that rested on their shoulders was condemned. In a case note, J.B.H.M. Simmelink considered that in light of the register’s increased reliability, a future status of a registration as ‘legal fiction’ should be considered. In 2015, this group was still very large. Interviews by the author with the RDW at the time confirmed that the RDW purposefully avoided to document the accounts of persons who called in (rather than put in an official request) with accounts of how they had been wrongly registered, as “this creates more space for people to avoid responsibility.” (Interview 28 November 2014, case notes in possession of author.)

567 MvT, Kamerstukken II, 1987-1988, 20 329, nr. 3, p.21 en 22.

568 In Falk and later judgements, the ECtHR accepted such administrative practices when other legal protections were in place, and, importantly, in light of the importance of actual road safety. EHRM, Falk vs. the Netherlands, 2003, 66273/01.

able to help. The administrative body has full discretion with regard to the Register.⁵⁶⁹ And in line with Administrative Law principles, all relevant information and interests can be made to count. The great number of ‘dehumanizing’ cases established because the RDW were reluctant to use this discretion over defending the ‘administrative truth’ that the Register represented. In what follows, the build-up of a sub-category of cases is illustrated. They are about ‘phantom vehicles.’ These have never been in possession or are no longer in possession of registrees.

How problems start for registrees

There are countless ways that kick off a registor's troubles. A shortlist on the basis of actual cases: documents get lost as a result of car accidents, theft, or sale. Institutions forget to issue records of transactions (demolition companies, car exporters.). Police refuse to register a theft, or forget to issue a document of demolition.⁵⁷⁰ Papers are taken in but not returned by an importing State's authority. Insurers forget to upload proof of insurance to the register the RDW uses as exclusive source of evidence. IDs, always required for requests, get lost or stolen, also unnoticed. Some people can't afford reissue fees, some forget to register theirs as stolen or can't get the police to register papers as such. Tax duties and fines are sent to the wrong address, when the single-authority's address register is incorrect or the RDW continues to use the ‘last known address.’

While registrees work to get their ‘assumed owner’ status changed, fines for non-compliance with all vehicle obligations⁵⁷¹ are issued automatically, periodically, and increase in several instances when they remain unpaid. This easily continues for many years and debts run well into tens of thousands of Euros.⁵⁷² To put a stop to this, registrees

569 “When the RDW assumes a person has ceased to be the owner, possessor or guardian of a registered vehicle it can terminate its entry in her name for it.” Art.40 b under 4/a Kentekenreglement.

570 When they do file a report, they also report the theft to the RDW, who suspends the vehicle duties. When vehicle is found, even as a total loss. The police report this back to RDW too and the RDW's compliance regime resumes. The registor is not notified of this however. The police (usually) contacts the person and asks them whether they want to reposess or let the police take care of demolition. When they decide to go for demolition, they need to ask for an official affirmation of demolition and send that to the RDW themselves, but this is badly known. An example of a case where this is at play is RB Amsterdam, 11-7--2014, zaaknr. AMS14/1111

571 Some registrees do re-insure their phantom vehicles, which is forbidden by law: insurers cannot insure an object without value. In the course of earlier research, the author conducted email exchanges with insurers to confront them with this fact. They confirmed by email that this problem was under their radar. (E-mail of Unigarant N.V. afdeling acceptatie, 21 August 2015, kenmerk # 4948559).

572 The severe disruptive consequences of this have been recorded Ombudsman reports, TV documentaries, news articles. See e.g. ‘Gijzeld door het Systeem. Onderzoek Nationale ombudsman over het gijzelen van mensen die boetes wel willen, maar niet kunnen betalen’; ‘De Monitor: Enorme boetes bij onverzekerde rijden’, *De Monitor*, 2015, <https://kro-ncrv.nl/persberichten/de-monitor-enorme-boetes-bij-onverzekerde-rijden>; ‘Een ongeluk komt nooit alleen: Een rapportage over een geslaagde interventie van de Nationale ombudsman naar aanleiding van een klacht over het Centraal Justitiele Incassobureau (CJIB) te Leeuwarden en de Dienst Wegverkeer (RDW) te Zoetermeer’ (Nationale Ombudsman, 13 January 2015); ‘Rechters laken gijzeling wanbetaler door justitie’, *AD.nl*, 24 February 2014, sec. Binnenland, <https://www.ad.nl/binnenland/rechters-laken-gijzeling-wanbetaler-door-justitie~aba57f1a/>; ‘Buitenhof’, *Buitenhof*, 12 April 2015, <https://tvblik.nl/buitenhof/12-April-2015>.

need to file Administrative review and appeal procedures against the RDW to end their registration; with the Public prosecutor to stop and annul the issued fines; and with the Tax Administration to stop and annul the road tax. If they are imprisoned for unpaid fines, they need to start a tort procedure against the State.

The fight to terminate registrations and annul issued fines

The standard RDW form to request a Register mutation has boxes for the requested documents, and no field to add additional information. When documents are missing, a negative administrative decision is issued. In this decision, the existence of an additional, official form for discontinuation requests on the basis of other grounds is not mentioned, nor is it explained (or acceded) that even without this additional form, the registree can submit other kinds of evidence such as pictures of scrapped or burnt-out cars, or witness statements, and the RDW needs to consider these.

When registrees somehow convince the RDW to strike their names, this is done on the basis of ‘courtesy’ since the RDW assumes no responsibility for the Registers real world accuracy. New fines will stop coming in, but past (years of) fines still hold. These can only be annulled on the basis of either a register ‘correction,’ i.e. a predated mutation to a moment before the starting date of the fines; or on the basis of courtesy of the Public Prosecutor and the Tax Administration; or when the RDW themselves take action and retracts the issued fines.

The last option was only discovered through previous research by the author. The RDW did not take the burden of the fines into account as an interest to balance in treating a request for retroactive mutation of the register, stating that the authority to issue fines was outside of their attributed competence. The fines were issued by the Central Legal Collection Agency (CJIB)⁵⁷³, who operate on behalf of the public prosecutor, and so the RDW referred to them to file a disproportionate hardship claim. Various rounds of FOI requests, and a series of exchanges with the CJIB however revealed that an RDW civil servant signs off, and sends off, batches of automatically generated fines (and not just assumptions of non-compliance) on behalf of the Public Prosecutor. The RDW was also attributed sole, and full, discretionary power to annul same fines, the Public Prosecutor may only halt collection. The CJIB are obliged to comply with a termination request of the RDW when a sole condition is met: the request needs to specify the reason for the requested annulment. A full second round of FOI requests was needed before the RDW conceded that this was indeed the procedure.⁵⁷⁴ In the documents that were eventually released, the RDW is stated to use only one reason:

⁵⁷³ Their own English translation, see <https://www.cjib.nl/en>.

⁵⁷⁴ The RDW initially responded that the described situation did not exist, and therefore they could not provide the requested reasons. RDW, letter of 5 August 2015, BZW.15.0353/0401.

“vehicle not in possession.”⁵⁷⁵ A surprising discovery in light of the RDW’s rationality thus far. This is explained below.

The RDW’s policy in statements of reasons

In the RDW’s system, the entry, mutation, and discontinuation of a vehicle constitutes a ‘register fact.’ Such facts are considered as building blocks of the ‘Register’s truth,’ or ‘administrative truth.’ And so, even if they would ‘believe’ a registree’s alternatively documented story of loss, they don’t want to predate a discontinuation in the register to align with the date of loss in the registree’s account, as this would falsify the administrative truth that is maintained by keeping to procedure.⁵⁷⁶ Genuine corrections are only granted in cases on the basis of particular evidence of identity theft.⁵⁷⁷ That option was added to policy in 2014, after the ECHtR in 2012 determined the State’s infringement on the family life of a person who accumulated 1753 cars in their name in the few days after their ID was stolen.⁵⁷⁸

In the statements of reasons of negative decisions about a correction request, the RDW typically argues how this corrodes the register’s purity. A more elaborate statement that the RDW submitted in a judicial procedure is more insightful. The case pertained a woman who had imported her car to Germany upon her immigration from The Netherlands. The German ‘sister authority’ took possession of the original car papers and issued a new license. They did not send the original license or a copy of the importing procedure to the RDW. The registree did not know something was amiss until many years later she visited The Netherlands with her young child. She was imprisoned at the Dutch border for unpaid fines and her child was placed in custody. All those years, fines for unmet vehicle registrations had been sent to the registree’s Dutch address. Although the RDW conceded they had received notice of emigration from the Municipality, the address was the last official address they had so they continued to use it. Forwarding post was not their responsibility.⁵⁷⁹ The judge sided with the registree. And as they were accustomed to do, the RDW appealed. “Reliability of the register is accomplished by protecting the administrative truth,” they argued, “which means that

575 Fax message from RDW/Unit Handhaving to CJIB, 16 April 2015, D424JS.

576 When I asked the department of legal affairs what the decision to offer courtesy and terminate a registration was based on if not belief of the registree’s story, the answer was that this was done on the basis of “a kind of belief.” Interview with the RDW, department of legal affairs of 28 November 2014.

577 Although strictly legally, the RDW Policy knows no restriction aside from the need that the situation is “exceptional”, which goes unqualified in the policy. Article 40 C, under 3, Kentelkenreglement.

578 A key argument of the court held that “the domestic authorities were no longer entitled to be unaware that whoever might have the applicant’s driving license in his or her possession was someone other than the applicant.” *Romet v. the Netherlands*, No. 7094/06 (ECtHR 14 February 2012); The ‘Kafka Brigade,’ A Dutch organization that investigates and helps organizations to fight ‘unnecessary bureaucracy’ A.C. Widlak, ‘Kan de overheid haar fouten corrigeren? #11’, *Stichting Kafkabrigade* (blog), 24 November 2018, <https://kafkabrigade.nl/home/publicaties/columns/-11-kan-de-overheid-haar-fouten-corrigeren-#idMxhU9NsiVhbrqHjaq0DtEQ>.

579 *Rechtbank Haarlem* 6 May 2015, HAA 14/2123.

a registry entry should always remain visible (...) this purity is apparently not very well understood, nor appreciated, by [registree]. That is understandable because this policy certainly creates far-reaching consequences for her. However, the trustworthiness of the entries for the sake of police investigation are enhanced.”⁵⁸⁰ And as was cited above, the RDW did not balance the fines and other consequences since those were stated to fall under other administrative bodies’ discretion. The high Administrative court ruling was in their favor.

The reasoning testifies to the RDW’s interpretation of the WAHV and the underlying VWV’s rationale of ‘road safety.’ Keeping known facts about actual cars outside of the register is certainly a contestable proposition. But in the range of measures to deal with cases of extreme hardship that were and are being developed (among them, ironically, an algorithmic scoring system to determine eligibility for the hardship status⁵⁸¹) none mean to intervene in the underlying system that produces the cases.

4.3 “A decision needs to be supported by a proper motivation”: explanation in the General Administrative Law Act (‘Awb’)

4.3.1 Introduction: scope and focus of the Awb’s explanation paradigm

This second part of the chapter discusses the Awb’s main explanation rules. Below, two important general characteristics of this explanation paradigm are explained: the restriction of what counts as decision to be justified, and the choice to allocate most ‘explanation output’ to the Internal Review level. The sections after discuss aspects of the historical codification process; the main objectives of the Awb’s explanation rules and critical perspectives on the paradigm and an experiment with ‘informal’ review procedures.

4.3.1.1 Dutch Administrative law’s restricted definition of ‘decision’

Civil servants are only obliged to provide reasons about those aspects and stages of a decisional process that fall under the Awb’s definition of an ‘order.’ Whether an act or uttering by an administrative employee, or of an entity’s presumed representative can be understood as an ‘order’ depends on a range of factors: the legal authority of who is ‘expressing decisional intent’; the specific attributed power that this person makes use of at that time; the subject of the act or uttering, whether it has been documented in writing, and whether the act or uttering pursues a specific legal consequence.⁵⁸² Orders

580 ABRvS 2015 03928, Hoger beroepschrift RDW 15 May 2015.

581 Stefan Kulk en Stijn Van Deursen, ‘Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek’ (Den Haag: WODC, 2020), 141 and further, <https://www.uu.nl/sites/default/files/tk-bijlage-onderzoeksrapport-juridische-aspecten-van-algoritmen-die-besluiten-nemen.pdf>.

582 Article 1.3 Awb.

come in two kinds: general orders (aimed at the wider population), and ‘decisions,’ aimed at individuals. It is these ‘decisions’ that the chapter is concerned with.⁵⁸³ Acts and utterings that do not qualify as decisions generally qualify as ‘factual actions.’ Such behaviors are regarded as the *carrying out* of policy, or of Awb orders themselves. Such acts are excluded from the Awb’s explanation regime, and cannot be legally objected to or appealed under the Awb’s process rules.⁵⁸⁴ A typical and topical simple example is when a person’s name acquires the label ‘potential fraudster’ through an administrative body’s investigative method or system. That occurrence itself, and the ensuing administrative decision to follow up on a notification (to investigate, or act on otherwise in preparation of decision making) do not count as Awb decisions. They do not need to be justified, or even mentioned. They are not open to review or appeal, such as when a notification proves to be false. Only an eventual decision can be appealed.⁵⁸⁵

The Awb does include a general complaints procedure that is open to grievances about non-decision behaviors. Administrative bodies are obliged to assess these complaints and issue a reasoned evaluation to the complainant. But the conclusion cannot be appealed. Unresolved complaints play out on the terrain of the national and municipal Ombudsman institutes. These analyze individual complaints, and the annual complaints reports that administrative bodies are obliged to provide them with. The Ombudsman institutes can and do also start investigations of their own accord. The institutes issue authoritative recommendations on the basis of their reports, but strictly legally, administrative bodies cannot be kept to comply with these.

This paradigm is criticized for several reasons. For one, automation developments have added, obscured and conflated steps in decisional processes so that the distinction can become practically impossible or unreasonable to make.⁵⁸⁶ There are also concerns about the obfuscation and the avoidance of accountability for legal effects that occur regardless of an eventual decision, such as the effects of investigations on personal, group or policy level.⁵⁸⁷ Another concern pertains to the inclination of administrative bodies to

583 In Dutch, the umbrella term is ‘Decision’, the individual variant is called ‘beschikking’. It is colloquially also referred to as ‘besluit’ or in English ‘order,’ especially with regard to explanation obligations since decisions fall under the explanation regime of ‘orders’.

584 Administrative bodies can be held accountable by suing the State in civil court.

585 The month that this manuscript was finalized, the Court of Amsterdam issued a decision that a screengrab of an automated, digital application form that terminated once a particular piece of information was entered upon which the screen said the applicant would not be eligible and was not allowed to apply further, was an appealable decision. It is unclear whether the Municipality will appeal, but when the decision becomes final, this has potentially major consequence for various other ‘unofficial’ decisions. ECLI:NL:RBAMS:2022:3066 Geautomatiseerde afwijzing na invullen online vragenlijst is bestuursrechtelijk besluit (Rechtbank Amsterdam October 2022).

586 Binns and Veale, ‘Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR’.

587 Investigations for example have legal effects when another administrative body needs an attest of good behaviour but this is suspended. ‘Fraudulous potential’ has also been ascribed to groups of people or whole neighborhoods, prompting policy level adjustments like increased surveillance (legally affecting privacy and equal treatment), in case leading to suspension of e.g. permits of benefits applications.

design for decisional steps that alleviate their bureaucratic burdens and or give them more discretionary space. The earlier example about the pre-eligibility process for material support is an example, as are the ‘clean house’ decisions that resulted from them. The information position of administrative bodies’ explainees is argued to have declined in the wake of the decentralization of the social security and care domains more broadly, with an instrumental role for the Awb’s decision-explanation paradigm.⁵⁸⁸

4.3.1.2 The internal review procedure: the entry (and exit) level for elaborate explanations

When a decision subject wants to object to a decision, their first step is to file for ‘review.’ These are internal procedures (although they can be outsourced) in which administrative bodies are allowed to repair shortcomings in their reasoning, or entirely change the ground of a decision (this will be discussed further in other sections.) Review outcomes (in the form of, again, decisions) can be appealed to: this takes the conflict to one of the Administrative courts, and possibly one of the Administrative High Courts eventually.

The Administrative internal review procedure is typified as ‘probably’ the most-used modality for conflict resolution in The Netherlands.⁵⁸⁹ A ‘lack of proper reasons’ in initial explanations are the ground of most internal review applications.⁵⁹⁰ The procedure became known as the main locus for the quality control of a decision, and the place where most reasons are made by administrative bodies in the first place. Initial statements of reasons (of individual decisions) are frequently reasoned less elaborately; as later sections will discuss, administrative bodies are for example allowed to refer to advisory reports, policy rules, or to ‘elaborate upon request’ in cases of positive eligibility decisions. Authors have argued that administrative bodies use the internal review procedure as a reductive instrument: to ‘elaborate upon request’ as a means to improve efficiency, not restricted to positive decisions.⁵⁹¹

In judicial appeal procedures, too, administrative bodies can repair their reasoning, or even substantively change them to the extent that an unchanged outcome rests on entirely new grounds. Some have argued this ‘denaturalizes’ the initial decision, and robs decision subjects of their original administrative due diligence and or due process

588 R.M. van Male, ‘Van motiveringscontrole naar bestuursrechtelijke rechtsvinding’, *Nederlands Tijdschrift voor Bestuursrecht* 2007; van Ettehoven en Marseille, ‘Afscheid van de klassieke procedure in het bestuursrecht?’.

589 In 2017, some 2 million reviews were counted. Marc Weber, ‘De bezwaarprocedure: Onderzoek naar verbanden tussen de inrichting van de procedure en de inhoudelijke kwaliteit van bezwaarbehandeling’, *Recht der Werkelijkheid* 38, nr. 2 (November 2017).

590 H.D. Tolsma, A.T. Marseille, en K.J. de Graaf, ‘Prettig Contact met de Overheid 5: Juridische kwaliteit van de informele aanpak beoordeeld’, Project Prettig Contact met de Overheid (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2013).

591 van Male, ‘Van motiveringscontrole naar bestuursrechtelijke rechtsvinding’.

rights.⁵⁹² The practice is considered to be problematic in other ways, too. The equality of arms that finds expression in how judicial procedures are open to citizens without legal representation is moot in absence of ‘inequality compensation.’⁵⁹³ Administrative bodies have the expertise, finances, and time to enter proceedings. For citizens, this costs—and although legal aid is financially covered for (only) citizens with very little financial means, decades of budget cuts, added eligibility requirements, and rises of obligatory court fees have hollowed out that system.⁵⁹⁴ In addition, decision subjects in precarious situations frequently lack the energy, oversight, time and overall bureaucratic tenure to engage in such procedures.⁵⁹⁵

But it is the character of judicial appeal procedures that adds most weight to the importance of internal review procedures, and with that, to the quality of administrative bodies’ justifications. Earlier in the chapter, it was cited that until very recently, courts only condemned the ‘reasonableness’ of administrative reasons if they were so blatantly inadequate that no reasoning was assumed to have taken place. Around the time of Awb enactment, the Judicial control of administrative decision making was improved and extended in other aspects. Judges gained investigative capacity. They could pick apart decisions and place them in additional (factual) light.⁵⁹⁶ But the discretionary space of the investigated administrative body remained largely intact, and their *qualification* of facts was still leading. In that sense, the Courts continued to perform only a ‘marginal test’ of administrative decisions: their ‘bare’ legality, the ‘mere’ reasons.⁵⁹⁷ Subjects who suffered very dire consequences of administrative decisions experienced that they were ‘finally heard’ by the presiding judges, only to be told that the administrative body’s reasoning was legally sound: typically phrased as (a variation on) ‘Administrative body X’s reasoning is not incomprehensible.’ But that judgment typically did not see to the *original* reasons; frequently these were altered several times already. In response to the feelings of deception that many subjects

592 R.J.N. Schlössels, ‘Kroniek beginselen van behoorlijk bestuur 2006’, *Nederlands Tijdschrift voor Bestuursrecht* 2006, nr. 6 (2006) under 2.8.

593 Marseille en Brenninkmeijer, ‘Een dialoog met de Raad van State na de toeslagenaffaire’, 604; A legislative proposal published at, or rather, just after, the time of writing called ‘strengthening the guarding role of the Awb’ includes a provision that allows citizens to submit additional evidence to prove their claims - but not to change their applications Ministerie van Algemene Zaken, ‘Memorie van Toelichting wetsvoorstel Wet versterking waarborgfunctie Awb (pre-consultatieversie 18 January 2023)’, ADD PAGE.

594 Tatjana Scheltema, ‘Wordt een advocaat slechts een privilege voor de rijken?’, *De Groene Amsterdammer*, 1 February 2019, <https://www.groene.nl/artikel/een-leemte-in-de-rechtshulp>.

595 ‘Weten is nog geen doen. Een realistisch perspectief op redzaamheid’; Ranchordás, ‘Empathy in the Digital Administrative State’.

596 T.J. Poppema, ‘Commentaar bij: Algemene wet bestuursrecht, Artikel 3:46 [Deugdelijke motivering]’, in *Encyclopedie Sociale Verzekeringen, Module Uitvoering sociale zekerheid en bestuursrecht* (Deventer: Kluwer,), last consulted 24 April 2019 referring to parliamentary papers MvT, PG Awb II, p. 463 and 175.

597 This is changing in 2022, partly as an effect of the failure of judicial protections of Benefits Scandal victims De Graaf en Marseille, ‘Exit willekeurstoets. Bestuursrechterlijke toetsing aan het evenredigheidsbeginsel na 2-2-’22’.

experienced, appeal proceedings were re-styled again. The focus became (an attempt to) resolve ‘the actual conflict,’⁵⁹⁸ rather than the conflict that was the subject of, or followed from, the initial decision. The change pushed administrative bodies’ original reasoning practices even further away from scrutiny, as the assessments focused on what administrative bodies brought to table in the court procedures even more. Van Male argued that this reduces the proceedings’ performativity with regard to positively influencing the practice of giving ‘proper reasons’ in the first instance.⁵⁹⁹ The remark adds to other concerns about judicial sensitivity to fundamental aspects of justice especially when these originate on other than individual levels: they are not allowed to check an administrative law for constitutional conflict (although they can test against treaty law), and an administrative body’s own rulemaking (e.g. policy rules) are excluded from appeals save exceptions.⁶⁰⁰ The former is currently under review and stands to be changed,⁶⁰¹ the latter is not.

4.3.2 The codification of the principle of Proper Motivation

4.3.2.1 “Just 9 (!) words”⁶⁰²: the “nihilistic” codification of the principle of motivation

Proper motivations of decisions are broadly seen as fundamental building blocks of trustworthy relations between Europe’s democratic governments and citizens.⁶⁰³ They inform the trust of citizens in the functioning of their (constitutional) democracies, including its controlling judiciary itself. So much was argued for in Chapter 2. This first section looks at the expression of these aims in the Dutch principle of motivation as it is understood to apply to the Administrative paradigm. Picking up from the introduction of legal principles in section 4.2.1.3, the principle, mostly referred to as that of ‘proper reasons’ has presence in the broader legal principles paradigm as well as a more specific appearance as a principle of proper administration. Among functions that are ascribed to the principle are its support and promotion of rational decision making, of justification, due process, and legal certainty.⁶⁰⁴ Proper reasons

598 And also raise more realistic expectations and better understanding of the procedure. Other aims were to shorten procedures and provide practically useful results. A.T. Marseille et al, ‘De praktijk van de Nieuwe zaaksbehandeling in het bestuursrecht’, ‘De Nieuwe Zaaksbehandeling’.

599 van Male, “Van motiveringscontrole naar bestuursrechtelijke rechtsvinding.”

600 Y.E. Schuurmans, ‘Toeslagenaffaire: outlier of symptoom van het systeem?’, *Rechtsgeleerd Magazijn Themis* 2021, nr. 6 : 206; Bröring en Tollenaar, ‘Menselijke maat in het bestuursrecht: afwijken van algemene regels’, 209.

601 ‘Eerste stap naar constitutionele toetsing door de rechter gezet’, Nieuwssite BZK, last consulted 8 July 2022, <https://www.nieuwsbzk.nl/2253818.aspx>.

602 Schlössels, ‘Kroniek beginselen van behoorlijk bestuur 2010’.

603 Council of Europe, *The Administration and You, A Handbook: Principles of administrative law concerning relations between individuals and public authorities*, as emphasized throughout. Grootelaar en van den Bos, ‘De Awb vanuit een procedurele rechtvaardigheids- perspectief: hulpmiddel, hinderpaal of handvat? Macht en tegenmacht in de netwerksamenleving’.

604 Schlössels en Zijlstra, *Bestuursrecht in de sociale rechtsstaat 1* section 8.3.7, under 57-60.

are expected to further the insightfulness of State decisions, allow for their review, promote their acceptability in case they come with necessary but adverse effects. In addition, a proper explanation regime promotes precise and just legal reasoning.⁶⁰⁵ This list combines formal and material dimensions; it sees to the governance of decisional procedures and to the quality of how such processes are justified.⁶⁰⁶

Schlössels, cited earlier to be in favor of precise, rather than abstract wording in the codification of principles in law, is critical of the Awb's codification of the "rich" principle of motivation. It is abstract, general, even "nihilist" in how it captures it in just nine words (in Dutch): "a decision needs to be supported by a proper motivation."⁶⁰⁷

4.3.2.2 *Can good decisions hide behind bad reasons (and should they be allowed to)? A salient codification discussion*

Before it was codified, the principle of proper motivation had developed through Administrative practice and case-law as a two-tier test of the 'properness' of a decisional process.⁶⁰⁸ Theoretically, the first tier sees to the establishment of facts; the second to the qualification of those facts in the process that leads up to the decision: the reasoning that was done to reach a conclusion. In practice, the tiers were not always clearly assessed separately. And in codification process, the causal link between the quality of a decision, and the quality of how it is reasoned came undone to some extent. This was preceded by discussions on whether the principle saw and should see to proper reasoning 'per se,' and if 'bad reasons' should necessarily disqualify a decision.⁶⁰⁹

The discussion, or at least the loosening of the bind can be understood to some extent by close-reading an authoritative 1984 report from the Commission on General Terms of Administrative Law (commissie ABAR.) It was a prominent advisory source for Awb legislators. In the ABAR report's description of the two-tier test as it functioned at that time, the first step as said meant to assess if facts were established in proper ways. The report adds that what facts are *relevant* depends on the legal and policy rules that govern a decisional process. These aims are interpreted by the administrative body, and whether they did so 'correctly' may only become apparent when step two is engaged with: the qualification of those facts. This second-tier assessment sees to the correct, just, and comprehensible explanation of (law & policy) rules and aims,

⁶⁰⁵ Beckers et al, *Motiveren, Over het motiveren van rechterlijke uitspraken*.

⁶⁰⁶ R.J.N. Schlössels, 'Kroniek beginselen van behoorlijk bestuur 2008', *Nederlands Tijdschrift voor Bestuursrecht* 2008, nr. 3 (2008).

⁶⁰⁷ Article 3:46. The law does list some additional obligations, but these not significant with regard to the explication of the principle. Schlössels, 'Kroniek beginselen van behoorlijk bestuur 2010' The law does list some additional obligations, but these not significant with regard to the explication of the principle.

⁶⁰⁸ 'Algemene bepalingen van administratief recht, 5e herziene editie' (Commissie ABAR, 1984), 136.

⁶⁰⁹ 'ABAR 1984', 143, 152, 170.

as expressed in the qualification of the factual account.⁶¹⁰ This is where ‘progressive uptake’ of developments in society are expected to promote legal development. E.g., a policy rule about spouse benefits was eventually made to count for cohabitant couples as well. After this became the norm, a decision that would *exclude* unmarried partners would be struck on the basis that it was ‘improperly reasoned.’ Another example named in the ABAR report was the application for broadcasting rights by the Humanist Federation. The body tasked with assessing their application had established the extent of their rights on the traditional basis of ‘membership count,’ but Humanists and their sympathizers eschew the term ‘membership’ because of the term’s religious connotations. On that basis a Judge ruled that the standard was therefore unfit, or rather, ‘improper’ to serve as a ground for the administrative decision, and the decision was struck. The terms used most in such cases was “the [administrative body/authority] could in all reasonableness not have come to this conclusion.”⁶¹¹

The report is however critical of what they saw as a *sloppy* use by the Courts of this qualification. It argues that Administrative Judges struck many administrative decisions for ‘a lack of proper reasons’ when they suspected material shortcomings in the decisional process, but lacked the time to properly investigate this.⁶¹² This meant, they argued, that perfectly defensible decisions were struck for how they were justified, and this shouldn’t happen. The authors argue that the demand of ‘proper reasons’ should principally function as ‘formal’ test. I.e., when a statement of reasons expresses material issues such as wrongful interpretation of an underlying rule, when it reveals unsuitable methods for establishing the facts, or when conclusions follow from an inexplicable or unjustifiable policy measure, the shortcomings can qualify as issues of form: as procedural demands that should not have negative consequences for the decision until (and if) it was established that the decision itself should have been different.⁶¹³

With the eye toward codification, they propose that Judges amend how they phrase their own statements of reasons accordingly. They should only use the term ‘lack of proper reasons’ when they are *sure* that material deficits exist, and that no reparation of reasons can make it better. They name situations of power abuse, of discrimination, arbitrariness, or eminent unreasonableness.⁶¹⁴ In other words, when a total lack of reason-*ability* exists. When this cannot be established, the judges should only declare that reasons are lacking in quality (that they are “not logically or rationally acceptable”) and that because of this, the decisional process needs clarification. Somewhat confusingly, they do concede that the ability to distinguish wrongs hidden

610 ‘ABAR 1984’, 167–71.

611 A term that has persisted through time until recently in the hardship clause test, as we saw.

612 ‘ABAR 1984’, 143.

613 ‘ABAR 1984’, 136–37.

614 ‘ABAR 1984’, 152.

behind ‘a lack of logic or ratio’ inevitably (also) depend on the insightfulness of the reasoned statement.⁶¹⁵

To conclude, the criticized contemporary practice of allowing administrative bodies very extended opportunities to repair reasons for decisions, are argued for in the report. The commission also seems to struggle with their advice that judges should only act, but still act, when they encounter the clearly wrongful situations. They foresee discussions on ‘the separation of powers’ when judges do find reasons to be unreasonable, since the problems in such cases may be the consequence of the lawmaker’s will.⁶¹⁶ In light of recent criticism *that* the Administrative judiciary shies away from justice-oriented reasoning, this deep-rooted cautiousness is interesting to note.

The ABAR advice was not adopted. But neither did the Awb codify a different definition, or envisioned relation, between formal and material dimensions of the principle of motivation. The Government considered how the two approaches would probably “conflate in practice.”⁶¹⁷

4.3.2.3 *Proper reasons: “it shouldn’t be much of a burden”*

The Parliamentary papers discuss how reasons need to be understandable enough to support the right of citizens to argue their case before a court without the help of a lawyer.⁶¹⁸ The ABAR Commission report included arguments on this point as well, and there it argues for quality that the government chose not to codify and that is found to be lacking in practice. The report argued that explanations need to fit the needs of a decision subject; reveal to them any underlying rules, the administrative body’s explanation of these rules, and the body’s interpretation of how the rules apply. An interpretation of a rule in primary law may also be explicated in a policy standard, in which case the demand for explainee understandability also applies to this standard. The report adds example questions for administrative bodies to answer in their a statement of reasons: what goals were aimed for, what interests were taken into account, what weights these interests were given, and why. It argues that “dependent variables” need to determine the length of the list of things that need explaining: the administrative body will need to explain how they assessed every factor they took into account, including personal information, external advice, mutual expectations, and alternative options.⁶¹⁹

615 ‘ABAR 1984’, 170.

616 ‘ABAR 1984’, 170.

617 MvT, Kamerstukken II 1988/89, 21 221, nr. 3, p. 108.

618 Grootelaar en van den Bos, ‘De Awb vanuit een procedurele rechtvaardigheids- perspectief: hulpmiddel, hinderpaal of handvat? Macht en tegenmacht in de netwerksamenleving’.

619 ‘ABAR 1984’, 138–39.

An earlier Awb draft did include more items of the list, but the explanatory memorandum of a later draft (which was amended again later on) argued that administrative bodies' explanation obligations should not amount to a burdensome task for them. Although they concede that explaining to subjects who are unfamiliar with applicable law and policy may require some additional effort, all in all, "[r]easons precede a decision in the common order of things," so when an administrative body struggles to provide proper reasons they may need to reconsider the grounds of their decisional authority.⁶²⁰ All in all, the law gave administrative bodies a lot of discretionary room to decide how far to take its reasoning. Whether they expected administrative bodies to make better use of it is unclear – but in light of the already long-term critique on explanation practices at the time,⁶²¹ they certainly had no reason to expect it.

4.3.3 The Awb's main explanation rules: objectives and critical perspectives

4.3.3.1 Knowable and insightful

The main explanation rules in the Awb are laid down in Title 3.7, "Reasons for decisions." The nine (Dutch) words are those of article 3:46: "a decision needs to be grounded on a proper motivation." In additional articles it is laid down that reasons need to mention legal decision grounds "where possible"; that reasons need to be communicated together with the decision unless urgency prevents to do so or unless it is "reasonable to assume" there is no need for reasons (see earlier: typically in case of a *positive* decision based on information that the applicant themselves provided.) In such cases reasons may be provided on an on-demand basis.

Contemporary Administrative Law handbooks qualify the provision to have formal and material dimensions, distributed over two tiers that need to be assessed: a decision's reasons must be *knowable* and *insightful*. 'Knowable' counts as the formal demand, and 'insightful' as the material. The insightfulness pertains to the need for a proper factual ground, and how the facts ground the decision: the actual reasoning.⁶²² The provision therewith requires to explain what facts and circumstances were considered in a person's case and how these were gathered, what general and individual interests were gathered and balanced (if any: depending on the discretionary space), and how their respective weights were determined. This is still not very explicit. Sometimes more guidance is established in case-law; it was for example established that when scientific information is part of the 'why' of decisions, this may need to be made insightful in individual cases such as when mathematical systems are used to calculate

620 MvT Kamerstukken 1988-89, 21221 nr. A, artikel 4.1.4 p. 65-67.

621 Filet, *Kortsluiting met de bureaucratie : over participatiemogelijkheden van burgers bij het openbaar bestuur*.

622 Tekst & Commentaar, Vakstudie Algemeen Deel, art. 3:46 Awb, aant. 1.6.1 De betekenis van het motiveringsbeginsel.

a decision subject's financial interests.⁶²³ But this is not knowable for explainees without professional assistance.

The preceding parts of the chapter included case descriptions that in themselves raise questions about the force and functioning of these explanation rules. It also cited various critical perspectives on the Dutch Administrative paradigm's legal and practical complexity, and explained the marginal influence of judicial scrutiny. This section relates several additional considerations to specific explanation norms, which helps to gauge the expressive and prescriptive 'justice' potential of the Awb's explanation rules later on in the chapter.

4.3.3.2 *Reasons in policy rules and the (lesser) motivational burden*

Administrative bodies may legally refer to other 'loci' of reasons, such as those laid down in policy *rules* and policy *guidelines*.⁶²⁴ Guidelines may only be referred to if they have been laid down in policy rules. If they are not, they need to be explained in the statement of reasons.

Policy rules are where administrative bodies make use of their secondary legislative powers. In other words, this is where quite some 'knowledge making' is done in the form of interpretation of the underlying primary laws. The Awb's explanation rules for decisions also apply to the making of policy rules as administrative decisions. But these justifications aren't submitted to individuals in meaningful ways: they are published (ideally, in practice not always – see earlier), but that does not mean there is reason to believe that citizens know about them – rather the opposite, since general bureaucratic literacy is established to be low.⁶²⁵ Neither are policy rules subject to parliamentary scrutiny – a problem in itself which was referred to earlier. The attributed power comes with its own 'hardship clause': policy rules are binding, unless special circumstances (of any interested party to the decision) make their application disproportionate in relation –this time– to the objectives of the policy rule.⁶²⁶ It was previously cited in this chapter how policy rules are used by administrative bodies to

623 In some subdomains, scientific standards and models are used particularly extensively. For example, decisions in the environmental domain inevitably rely on such methods. As Van der Veen discusses, codification afforded public institutions some handles for regulating and assessing such standards, but clear higher-order legal guidance is absent, and administrative bodies will need to argue and defend its chosen methods. Gerrit van der Veen, 'Digitalisering in het omgevingsrecht en mogelijke invloed op de Awb: De burger tussen de ambities en doelstellingen van de Awb', in *25 jaar Awb: in eenheid en verscheidenheid*, edited by A.T Marseille e.a. (Deventer: Wolters Kluwer, 2019).

624 Article 4:82 Awb. Left undiscussed are reasons in advisory reports: when an administrative body adopts conclusions of an official, in casu mandated advisory report, the statement of reasons may refer to the report when it has been made available to the explainee (art. 3:49). Not so when the departs from the report's conclusions, though: this will need to be reasoned. (art. 3:50).

625 See earlier remarks on low average legal/policy literacy and the complexity of the administrative domain in particular.

626 Article 4:84 Awb.

reduce their burdens of explanation to individuals.⁶²⁷ Up to a 2016 High Administrative Court ruling, administrative bodies also did not need to *explain* a negative decision about a subjects' claim to special circumstances. They were allowed to declare that the subject's situation was foreseen in the policy making, and accounted for in the policy rule. This reduced the investigative, as well as the motivational burden on administrative bodies.⁶²⁸

After the ruling, this 'explanation of the explanation rules' was no longer deemed acceptable. Administrative bodies are now obliged to establish (and declare) that there *aren't* any special circumstances that would *prohibit* the application of a policy rule.⁶²⁹ But they only need to explain how they qualified a subjects' circumstances when that person objects to this decision.⁶³⁰ At the time of these developments, lawyer and legal scholar Franssen speculated that administrative bodies' statements of reasons would start to improve in quality—if only because administrative bodies would make fewer policy rules now that such rules' main value was reduced: a lesser motivational burden.⁶³¹

The reliance on the preparedness and ability of decision subjects to object to the decision is still problematic however. As was argued at various points in the preceding parts of the chapter, that can amount to a significant and arguably unfair burden. Taken together with the existing tradition to reason initial statements less elaborately, and taking into account the Council of State's advocacy for the review procedure as ADM explanation level *par excellence*, the question can be raised what meaning the 'completeness and understandability of reasons' is set up to have for the relationship of first explainer and explainee. The next section further engages with this concern.

627 P.E.M. Franssen, 'Beleidsregels en de inherente afwijkingsbevoegdheid: de nieuwe lijn van de Afdeling.', *Praktisch Bestuursrecht* 2017, nr. 7 (13 December 2017).

628 Franssen.

629 R.J.N. Schlössels, 'Kroniek beginselen van behoorlijk bestuur 2000', *Nederlands Tijdschrift voor Bestuursrecht* 2000, nr. 9 (2000).

630 Franssen, 'Beleidsregels en de inherente afwijkingsbevoegdheid: de nieuwe lijn van de Afdeling.'

631 Franssen; But, literature warns to never underestimate the Administrations standardization tendencies. Immigration lawyer Peeters' investigation into the reasoning of a large series of hardship cases revealed policy where none was known. Years of research and FOIA requests revealed how a hardship clause had been unknowingly, therewith illegitimately explicated with a list of weights that determined the value of various hardship factors. Inspiration for the initiative seemed to have come from the highest political level: the Immigration Minister at the time had circulated a questionnaire among Mayors, to help them assess an applicant's chances of eligibility before submitting a personal hardship appeal to him. (At the time of writing, this Ministerial discretion has been discontinued.) These weights were never mentioned in statements of reasons. M.J.M. Peeters, 'Hoe wordt de discretionaire bevoegdheid in schrijnende situaties gebruikt?', *A&MR* 2018, nr. 1 : 7.

4.3.3.3 External insightfulness v. a relational understanding of ‘proper’ reasons

It is understood that the statement of reasons needs to provide the kind of insight that lets external parties assess compliance with the *principle* of motivation, due process, and other applicable principles.⁶³² In other words, it needs to serve Administrative Law experts, as these are the ones who are expected to have this capability. Think of an administrative body’s internal review department, decision subjects’ lawyers, the Judiciary, but also the Ombudsman. Not included in this list are explainees, themselves. Arguably, decision subjects’ responsible interaction with the State *should* not depend on their having legal administrative expertise. But since the Awb’s explanation rules themselves don’t include a more comprehensive account of the expected reach and quality of statements of reasons, explainees are made dependent on the (availability of) expertise of others to quite a large extent. And such expertise is not granted to them in Administrative Law itself. The idea, as was explained, is that no-one should need an expert at their side to engage with the State’s reasoning in their case.

This can be signaled as problematic in light of the more extensive rights that subjects have on the basis of a principled understanding of Administrative rules. It was for example explained that all information, and all interests, should be allowed to play a role *in principle*. This is especially pertinent since there is an unclear burden of proof with regard to the establishment of facts: Administrative Law lacks a bespoke regime for evidence. E.g., when policy rules restrict the information and interests that ‘count’ (like in the phantom vehicle cases), and no-one explains a subject that they can still argue their case on the basis of other evidence and circumstances, they have no reason to bring such information to the table.

These choices of the Awb are of interest to note in light of the earlier cited 1970s research, and of pursuant and persisting critique on the quality of State-citizen interactions. It was and is known that citizens’ understanding of their decisional processes needs to be catered to *during* the whole process. It should neither start nor stop with the statement of reasons, especially if their ‘bureaucratic literacy’ is low to begin with.⁶³³ There are myriad literatures for lawmakers to engage with with regard to the design of explanatory exchanges,⁶³⁴ and some modalities are indeed experimented with, as will be discussed in the last section. But these, again, are aimed at the review phase. In 2019, then National Ombudsman Brenninkmeijer voiced his critique on what he called the ‘legalistic’ application of

632 Rens Koenraad, ‘Op zoek naar algemene beginselen van behoorlijk Burgerschap in het Nederlands bestuursrecht’, in *25 jaar Awb: in eenheid en verscheidenheid*, edited by A.T Marseille e.a. (Deventer: Wolters Kluwer, 2019).

633 Filet, *Kortsluiting met de bureaucratie : over participatiemogelijkheden van burgers bij het openbaar bestuur*.

634 Contemporary procedural justice research for example discusses how the modalities of explanatory exchanges matter, such as that face-to-face exchanges are beneficial. Grootelaar en van den Bos, ‘De Awb vanuit een procedurele rechtvaardigheids- perspectief: hulpmiddel, hinderpaal of handvat? Macht en tegenmacht in de netwerksamenleving’.

statements of reasons’ demand for ‘know-ability.’ He argued that the statements need to be *reasonably understandable*, rather than *understandably reasoned* in abstract terms. An improved interpretation of the principle of proper motivation should explain ‘proper’ in “relational” terms: the relation of explainers and explainees.⁶³⁵

4.3.3.4 Decisions ‘as’ arguments v. decisions ‘and’ arguments

This section picks up on the Awb codification discussion about the purported significance of a badly reasoned decision. The ABAR report suggested that motivations of low quality should initially be assessed on their own merits. Bad reasons don’t necessarily represent bad decisions, and so administrative bodies should be allowed to repair them. This indeed became regular practice.

The understanding of legal reasoning that is expressed in this is arguably problematic. The point of giving reasons is to ‘tell’ on how a conclusion was reached.⁶³⁶ As per the same ABAR report, this includes an administrative body’s interpretations of the laws that their authority is derived from. In other words, the point is to ‘get inside the head’ of the decision maker, and engage with the quality of their argument. If that quality is poor, and cannot ground the decision, the official response to that expresses the weight (and type of weight) that is ascribed to ‘giving proper reasons.’

This section will not again engage with confusions that arise from all too literal interpretations of what ‘inside the head’ means. But it should be iterated that debate exists about the value of explainers’ ‘testimonies’ because giving reasons is a linguistic act. Explainers can lie, and are purported to lack insight into their own decisional processes for neurological and psychological reasons. This thesis, as was made clear, does not agree that these are obstacles to meaningful explanation. But that does not mean that there are no merits to problematizing the ‘language of reasons’ to assess the justice of decisions. Language itself is an established source of representational injustices, as also becomes apparent in AI’s language models.⁶³⁷ Language can certainly also be *used* unfairly to deliver judgments: from skewed arguments to the use of complex ‘legalese.’⁶³⁸

635 Alex Brenninkmeijer, ‘De burger tussen de ambities en doelstellingen van de Awb’, in *25 jaar Awb: in eenheid en verscheidenheid*, edited by A.T Marseille e.a. (Deventer: Wolters Kluwer, 2019), 44.

636 As Schuurmans explains, in law, the space of acceptable, proper reasoning is ruled by the “conviction raisonnée” Schuurmans, “Bewijslastverdeling in het bestuursrecht,” 16.

637 Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan, ‘Semantics derived automatically from language corpora contain human-like biases’, *Science* 356, nr. 6334 (14 April 2017): 183–86.

638 In an effort to right this particular wrong, New Zealand recently passed the “Plain Language Act.” MP Rachel Boyack was cited in *The Guardian* to say “People living in New Zealand have a right to understand what the government is asking them to do, and what their rights are, what they’re entitled to from government.” Tess McClure, ‘New Zealand Passes Plain Language Bill to Jettison Jargon’, *The Guardian*, 19 October 2022, sec. World news, <https://www.theguardian.com/world/2022/oct/20/new-zealand-passes-plain-language-bill-to-jettison-jargon>.

But that is precisely why the quality and understandability of the language of reasons must be accounted for. Language is also the instrument with which to object to reasons. In ruled explanation paradigms, language is the tool with which to contest the decisional process that a set of reasons represents.⁶³⁹ For this, reasons need to give sufficient insight.

4.3.3.5 Case illustration 3: experiments with informal review procedures

This case illustration discusses an experiment that was part of the Ministry of the Interior's 'pleasant government contact' initiative. The initiative started in 2007. Its goal was to create a platform for developing 'shared ideals of a more horizontal, responsive approach to citizen relations, with a focus on contact and communication.'⁶⁴⁰ The platform supports administrative bodies with the implementation of less 'procedural' ways to deal with citizens' complaints and objections. It notes that 'a lack of proper reasons' is one of the main grounds for review requests,⁶⁴¹ and meant to address concerns that the Awb review procedure's affordances to engage in practical, problem-solving citizen relations had failed to materialize.⁶⁴²

After several years of trialing 'informal approaches' to build expertise and experience, a 2013 booklet from the Ministry reports on three successful 'treatment modalities.'⁶⁴³ The first is 'a sympathetic ear, clarification and explanation';⁶⁴⁴ the second, 'review/adaptation of decisions'; and the third, 'creative solution.' The modalities hang together, and notions with regard to explanation are part of all three. The experiments recount of various (possible) improvements to situations about which critique was cited in earlier sections. But on the basis of the studied documentation it is unclear whether the deeper origins and causes of the improvable State relations are investigated, and engaged with. In addition, the juxtaposition of (formal, strict) proceduralness and responsiveness and informality arguably produced a problematic understanding

639 As Hildebrandt writes, contained in the 'Rule of Law' are "individual rights that enable legal subjects to speak law to power" Mireille Hildebrandt, 'Law As an Affordance: The Devil Is in the Vanishing Point(s)', *Critical Analysis of Law* 4, nr. 1 (2017): 119.

640 This description has since been replaced on the project website, which is now focusing almost entirely on 'informal approaches' for administrative bodies. The site's title is still 'pleasant contact,' but on the site it now says "Responsive Government / appropriate contact." <https://www.pcmo.nl/index.php/>.

641 Tolsma, Marseille, and de Graaf, 'Prettig Contact met de Overheid 5: Juridische kwaliteit van de informele aanpak beoordeeld'.

642 Ministerie van Binnenlandse Zaken, 'Kamerbrief met Kabinetsreactie "Hoe hoort het eigenlijk? Passend contact tussen overheid en burger"', 9 March 2016; Marc Wever, 'Bezwaarbehandeling door de overheid anno 2016: Vooral vernieuwing op papier?', *Nederlands Juristenblad* 2016, nr. 44.

643 Tolsma, Marseille, and de Graaf, 'Prettig Contact met de Overheid 5: Juridische kwaliteit van de informele aanpak beoordeeld'.

644 'Toelichting' and 'uitleg' in Dutch both translate to 'explanation.' 'Toelichting' is considered to be the less in-depth variation, comparable with a clarifying pop-up instruction in fields of online forms, where explanation is more akin to elaboration.

of what proper explanations are for, raising questions of legitimacy. Both issues are explained below.

To start with the first issue: in the booklet, various municipal research reports about their own ‘informal review’ trials are referenced. One reports how municipalities increasingly engage in mediation or pre-mediation with decision subjects who wish to object, adding how “some [municipalities] refer to this as an informal approach.” The goal, they state, is the same: to come to a solution without a legal procedure.⁶⁴⁵ Success seems to be measured by retracted objection procedures. Another project reported that review case workers investigated what help-seeking reasons existed “behind” an applicant’s denied eligibility request (under the earlier discussed Social Support Act), in addition to the traditional rule-based review assessment.⁶⁴⁶ This helped to find alternative, tailor made solutions. Both descriptions show a departure from the critiqued ‘narrow competence’ paradigm that prohibits to take facts and interests into account that are not immediately relevant to the law or domain that a support request is connected to. A broad view of possibly applicable laws is promoted, bringing into view a range of legal grounds to choose from, to meet the help request. A similar suggestion was later voiced by the national Government: specialized administrative bodies are encouraged to cooperate with each other to support such investigations.⁶⁴⁷ This ‘broad competence’ mindset is (also) advocated by the Institute for Public Values whose clients include various Government and municipal bodies. In addition to engaging with more laws, they argue that civil servants need to engage with applicable laws on deeper levels: to engage with their rationales and not just with the provisions these were codified into. Starting from the premise that the General Principles of Proper Administration are very badly adhered to in the social security and care domains, they argue to adopt a more productive set of “Principles of Proper Customization.”⁶⁴⁸ Among other things, these principles promote that administrative bodies co-create detailed, well argued, and properly discussed support plans with citizens.

But the question can be asked how useful it is to establish new principles if principles that are argued to pursue the same thing aren’t made use of well enough. The review procedure as described in the Awb already speaks of the need to investigate ‘alternative options.’ And the elaboration, explication, and/or further codification of the principles of due diligence and proper motivation are already expected to make the necessary difference. The questions also relate back to the discussion on the ‘character’ of Administrative Law: whether the Awb should be understood as a legal expression of

645 De Koster, ‘Wisselend succes na mediation sociaal domein’.

646 Frits De Jong, “‘Er is een Wabo nodig voor het sociaal domein’ | iBestuur”, last consulted 17 January 2019, <https://ibestuur.nl/partner-vng-realisatie/er-is-een-wabo-nodig-voor-het-sociaal-domein>.

647 ‘Brief van de ministers voor Rechtsbescherming en van Binnenlandse Zaken en Koninkrijksrelaties’ (Parliamentary papers 34 775 VI, no 4, 22 January 2018), 8.

648 Harry Kruijer, ‘De Algemene Beginselen van Behoorlijk Maatwerk’, *Instituut Publieke Waarden* (blog), 1 December 2016, <https://publiekewaarden.nl/de-algemene-beginselen-van-behoorlijk-maatwerk/>.

public authority, or an authoritative expression of principled justice, that expects civil servants to make use of their principled, discretionary space.⁶⁴⁹ When civil servants are given more discretionary space on a principled basis, as the experiments seem to suggest (since they don't argue to change the Awb itself), then the expectation that they will use that space and use it to the beneficence of explainees needs to be grounded. That ground is lacking.

To compare, the National Ombudsman argued how their “principles of proper administrative behavior” express demands of human interaction beyond what can be expected from demands rooted in ‘legality’, in other words, in law. This view is criticized by Schlössels for how it waters down what we *should* understand legality to mean, especially where ‘informality’ is experimented with, or even becomes the norm.⁶⁵⁰ This bridges to the second issue.

As was cited above, municipalities that participated in the informal review experiments sometimes described their informal approaches as ‘mediation.’ A surprising term to use for the conflict resolution between two fundamentally unequal parties, the State and an individual citizen. Whether or not this notion of horizontal equality watered down administrative bodies’ awareness of demands of legitimacy cannot be said, but that legitimacy itself was not served well enough is apparent. For example, municipalities reported how more and better explanations led many subjects to retract their decision review requests,⁶⁵¹ but since ‘better explanations’ are not necessarily put to paper, it will be hard to assess whether justice was done. Some municipalities also encouraged citizens to contact their case workers before objecting to a decision, with the aim of avoiding a review procedure altogether.⁶⁵² But these conversations as well are not recorded or reported in case files, and cannot be studied. Some municipalities were revealed to have actively frustrated the engagement of subjects’ legal representatives by ‘informally’ approaching subjects directly.⁶⁵³

Not all review case workers felt at ease with the less clearly governed discretionary space, and were reportedly hesitant to use it.⁶⁵⁴ A lack of training and experience was said to be involved in this too: an ‘informal’ telephone call is a different and more personal experience than a formal hearing with the legal department. One research

649 van den Berge, ‘Bestuursrecht na de toeslagenaffaire’.

650 Raymond Schlössels, ‘De Harde Kern van Behoorlijkheid: Over rechtmatigheid, behoorlijkheid en de Nationale ombudsman’, Nijmegen Migration Law Working Papers Series, 2014, nr. 3 (2014).

651 De Koster, ‘Wisselend succes na mediation sociaal domein’.

652 Wever, ‘Bezwaarbehandeling door de overheid anno 2016: Vooral vernieuwing op papier?’, 3432.

653 The cited municipality names practical objectives like location or the costly time of reps, but personal communications from other municipalities told (me) another story: lawyers were seen to create a highly rights-based, adversarial stance with complainants. Wever, ‘Bezwaarbehandeling door de overheid anno 2016: Vooral vernieuwing op papier?’

654 A.G. Mein en Bert Marseille, ‘Informeel aanpak bij bezwaar: rapportage werkpakket 2: de belfase’ (Amsterdam University of Applied Sciences, AKMI, 2019), 36.

team who tracked the experiments experienced that although administrative bodies' team leaders were enthusiastic about researchers sitting in on the 'informal' calls, case workers themselves tended to avoid their presence by scheduling calls when the researchers were absent. The researchers were left to study the case files, but soon found out that these held little notes on what was discussed and decided in the informal review process.⁶⁵⁵

The illustrations attest to review procedures whose legitimacy cannot not be assessed. The government acknowledged these risks.⁶⁵⁶ In a report on research findings, they sketch a "worst case scenario" in which "the review department and the review secretary both hold the incorrect conviction that the decision under scrutiny is legitimate, and the secretary uses his communicative skills to include the decision subject in this belief." They argue that they have no reason to believe such things will actually happen, but that on the basis of their analysis of informal review cases in which initial decisions were upheld, they also cannot conclude the decisions to uphold were legitimate.

This part of the chapter ends with one last citation of Filet's 1974 report. It argues that although the complexity of Administrative law is an obstacle to meaningful explanation practices, the solution to this is not to create less detailed law policy—rather the opposite. The book argues for more clearly defined rules and procedures. In a contemporary critique on the informal approach, Damen advises that civil servants should be trained to *not* give 'slapdash' explanations of complex (procedural) rules in informal communications, and to record explanation conversations as a policy.⁶⁵⁷ In a parallel critique, the National Ombudsman himself voiced legitimacy concerns with regard to administrative bodies' informal complaints procedures, warning that legal safeguards need to be observed, such as written confirmations of commitments and agreements, and properly reasoned decisions.⁶⁵⁸

655 The researchers also described legally trained review case workers were described as to (roughly) fall into two categories: 'legal-autonomous,' who grappled with their 'unbounded' freedom, and 'responsive,' who did not. *Mein en Marseille*, 39 and personal communication with researcher.

656 Tolsma, Marseille, and de Graaf, 'Prettig Contact met de Overheid 5: Juridische kwaliteit van de informele aanpak beoordeeld'.

657 Damen, 'De autonome Awbmen?', 634.

658 'Terug aan tafel, samen de klacht oplossen: Behandeling klachten over zorg, jeugdhulp en begeleiding naar werk', 9.

4.4. The Awb's explanation governance in terms of the modeled duties of explanation care

4.4.1 Introduction: analysis and reporting structure

The first part of the chapter (the 'what, who, how') described the attributed knowledge and (individual) decision making powers of the Administrative domain, expectations with regard to administrative bodies' functioning in political systems, and several relevant Administrative process rules that govern these capabilities. It included a brief description of civil servants as a type of decision maker, focusing on the much-described tension between their obligation to apply and abide by State rules, while they are also expected to resist oppressive State politics. The second part of the chapter described the domain's main explanation rules. In all these descriptions, acknowledged and apparent tensions, and scholarly critique from within the field about the fields' rules and functioning was engaged with.

The third part of the chapter relates these two parts through an analysis of them at the hand of the modeled duties of explanation care. The Model is a categorization of fundamental explanation values that deserve and require institutional expression and application. The analysis of the findings from the first two parts will be structured as follows. Each element is treated in a separate section. Each of these sections starts with the text of the element as it was defined in Chapter 3, including its more elaborate description. After this, pertinence and recognition of the elements' aims and values are discussed: how they 'make sense' in and for the field, and were engaged with in literature about it. These findings are related to the expression of the aims and values (or lack thereof) in the domain's explanation rules.

With this, a critical description of the domain's legal explanation paradigm is constructed. The elements' descriptions serve to answer (in part) the third research question. They will not be further condensed or concluded about; they will be parked to be instrumentalized in Chapter 6. That chapter draws and combines lessons from both domains' analyses: lessons that need to inform the (further) development of ruled explanation paradigms in AI-informed times.

To come to this part's analysis, findings of the preceding sections were hand-coded for Model element relevance. While discussing them in the element sections, several overarching themes appeared that repeated under most elements. For example, the theme 'discretionary space of Administrative decision makers' included findings with regard to the investigation of underlying knowledges (element one); about what explainee facts and interests can legally be made to count (element two), with regard to breadth and depth of justification (element three), and was also apparent with regard to what is or is not recorded about explanation (element four). The other themes were 'information positions of decision subjects'; 'what should reasons represent?'; and

‘reasons on demand/reliance on review.’ The themes are used as categorizing headers in the element sections.

4.4.2 Element one: investigating explainer authority

Explainers are obliged to investigate their own social-epistemic positions with regard to their decision-making modalities, and their domain’s underlying (input) knowledges in order to assess their role (=explainer) authority: does the explainers’ understanding justify their authoritative and trustworthy explainer position? If no (or can’t investigate), rebel.

This element obliges that explainers avoid to become an instrument of unjust (‘bad’, oppressive) knowledge practices, and are able to explain their ‘avoidance strategies’ to their explainees. To what extent they need to in fact explain these strategies is best determined in a decision domain’s context. More positively expressed, this element promotes that explainers are able to communicate how, and not just that they are trustworthy ‘knowledge practitioners,’ and not just accountable decision makers. The point at this stage is to link the self-reflection of explainers to their position of authority *vis-à-vis* explainees, as part of responsible practice. The need for explainers to rebel exists when explainers feel incapable to do this, for example because they don’t have access to justificatory sources or aren’t afforded the time, or means, to investigate.

4.4.2.1 General recognition for element one: expectations with regard to civil servants’ critical engagement

To reiterate, the Benefits Scandal fore fronted concerns with regard to the ‘moral apparatus’ of the Administration’s civil servants.⁶⁵⁹ But discussions about what made them do what they did in the Scandal are inconclusive on important points. Such as, whether they were expected to execute the underlying law in the strictest possible way, or whether they did so of their own accord. Whether the discriminatory effect was rooted in law already, or entered in administrative policy. There is also discussion about the legal means that were available to them to *avoid* to do what they did: about whether they could have, or should have, balanced the proportionality of adverse effects, and more fundamentally reflected on their duties as actors in the *trias politica* of a constitutional democracy.⁶⁶⁰

⁶⁵⁹ Mainly the Tax Administration, but others as well. The Tax Administration for example presented unfounded ‘suspicion of fraud’ labels to municipalities about ‘clients’ of their local administrative bodies. And administrative bodies’ child safety teams outplaced a large amount of children from families whose lives were disrupted by the policies. All these administrative bodies are asked to improve their own moral radars and practices.

⁶⁶⁰ Sections 1.3, 1.2.1.2 and throughout.

The consulted literatures attest to how such discussions are fundamental to, and for, the domain. Administrations are no longer regarded as executive branches only, but relied on to resist oppressive political influence and act in ways that lets decision subjects and others hold them to account if they don't. Concerns about 'unthinking,' and deliberately cooperating and collaborating bureaucratic forces were loud after the Second World War. And although descriptions of innate tensions with regard to their tasks persist (the prototypical rule-abiding, impersonal Weberian bureaucrat v. the personally engaged and responsive civil servant⁶⁶¹), the establishment of the human rights regime and their embedding in national constitutions is still expected to hold Governments' executive forces to the right kind of values and inform the conscience that all *trias* members are relied on to engage in their work.

The need to govern administrative bodies on the basis of theoretical but also realistic expectations of non-oppressive treatment are pertinent in light of the knowledge and decision-making powers attributed to them. Before individual decisions come into play, they have an active role in guarding against unjust treatment in how they use their secondary legislative powers. Put differently, they also need to guard themselves (against) their own political tendencies.⁶⁶² In light of ADM developments, this raises specific concerns of the Council of State. Administrative bodies, they argue, are where law translates into (automated) policy, and so, the civil servants in place become the 'most knowledgeable' person with regard to policy and how individual decisions based upon it are made.⁶⁶³ And since even 'simple' automated methods used in policy have led to dire consequences for individuals they argue that the regulation of their duties under the principles of due process and justification needs to be improved. Not so the Government, who argued that administrative bodies were perfectly capable of designing and executing the cited policies in more justice-serving ways under the current regime. But the longevity, as well as the institutional and disciplinary diversity of concerns about the lack of properly justified and humane decision practices warn to govern the domain on the basis of (more) realistic expectations.⁶⁶⁴ This is especially true in light of the 'harsh political climates' alive in many European countries where the post WWII welfare state became discredited. The parallel development of hardened asylum policies is seen to have added ethnic color to the 'othering' of whoever is in need of State support.⁶⁶⁵ The next sections briefly engage with some findings that attest to the strengths and weaknesses of the current regime in terms of element 1, which helps to inform these realistic expectations.

⁶⁶¹ Section 4.4.2.

⁶⁶² Section 4.2.1.2.

⁶⁶³ How this abstracts from the fact that such knowledgeability has rather shifted to the tech companies that provide the automation will be left aside for now.

⁶⁶⁴ Sections 4.2.3.5-4.2.3.6 and throughout.

⁶⁶⁵ Section 4.2.1.2.

4.4.2.2 *Discretionary space for decision makers: what guidance for principled engagement?*

In consulted literatures, the discretionary space of administrative bodies with regard to their policy and individual decision making is cited by some to be too restrictive, by others as too broad. This reproduces in an unhelpful dichotomy between notions of ‘formality’ and ‘informality,’ where formality is cited as counter-productive of civil servants’ humane engagement but proposals of enlarging an ‘informal,’ less strictly regulated space is cited as risky.⁶⁶⁶ This raises questions about the Awb’s instructions about the *use of* the required discretionary space: what values civil servants are *formally* meant to engage with in absence of minute instructions. Put differently, these are questions about why discretionary space produces problems—but one can also wonder why more discretion should be understood as, and governed as, ‘informal’ to begin with. These questions are further discussed below.

To start with arguments about perceived restrictions: Dutch Administrative law is known for its high level of complexity and detail. Consecutive governments have promoted strict compliance regimes, as well as the implementation of standardization and automation for carrying out policies, and methods to check compliance with them. All these factors are seen to reinforce each other. Specializations and competences are distributed over administrative bodies, whose bespoke information establishment and management complicate the interoperability necessary for individual and especially ‘tailor made’ decision making. Civil servants lack the necessary insight and oversight, as well as the practical means to act more responsively and therewith more humanely. These dynamics are seen to corrode the space for reflection and ‘value rationality’ of civil servants.⁶⁶⁷

But whether reflection and value-rationality *naturally* blossom with more discretionary space is questionable, depending on the kind of engagement that such arguments mean to refer to: with values in primary laws and (their expression of) the political climates in which they are enacted; with codified legal principles, other legal principles and individual human rights-oriented justice; or with an alertness to institutional racism and discrimination, and other less apparent forms of ‘state oppression.’ The latter kind of engagement arguably requires the most independent type of ‘moral conscientiousness’ and was found lacking on a grand scale in the Benefits Scandal. Engagement with values ‘in the middle,’ legal principles and human rights-derived values have not prevented the grand scale injustices either. The former, sensitivity to values that (explicitly or implicitly) express in primary laws, does not seem to need much encouragement: already strict regimes were given exacerbated effect in policies enacted by administrative bodies. Various examples from the domain testify to this.

⁶⁶⁶ Section 4.3.3.5

⁶⁶⁷ Section 4.3.3.5

For example, the choice to exclude citizens with vehicles registered in their names from the Criminal Law regime and its presumption of innocence (in ‘minor’ misdemeanor cases) was made in primary law, but the reliance on administrative bodies to enact a sufficiently protective regime was moot. Even after an adverse ECtHR ruling, and even after local Judges had started to refuse to effectuate Administrative punishments, large numbers of persons were socioeconomically run-down by the RDW’s and related administrative bodies’ unchanged policies, and procedural routes that could help them were actively obscured.⁶⁶⁸ The same reliance on administrative bodies was part of the decentralization of social security and care domains. These laws came with very large measures of discretion. But the Participation Law’s embedded wrongful ideology (i.e. calculative, self-sufficient and bureaucratically capable model citizens) and its inclusion of harsh punishment affordances in case of non-compliance resulted in harsh and distrustful policies—and in unjustified and unsightful decisional practices.⁶⁶⁹

In both cases, the response of administrative bodies to the controlling judiciary is of interest to note. In the phantom vehicles cases, judgments that ruled in favor of victims were either appealed or interpreted as narrowly as possible, including an ECtHR ruling. And in the Participation law cases, effects of the condemning judges of one of the High Administrative courts were repeatedly ignored. These considerations are important in light of the calls for the further development of the principles of motivation and due process, and the principle of ‘meaningful contact with the State’ that is currently being developed. Some authors argued that eventually, protection from Administrative power abuse should be located outside of the Administrative legal paradigm. But the case illustrations, including the response to authoritative, rights-based corrections from outside (when they come—the relative engagement of the judiciary itself is another issue) arguably promote to strengthen the right kind of ‘value rationality’ in the most primary instructions of civil servants. Not as ‘informal’ beneficence, but as formal demands.

This seems especially pertinent since policy rules (salient places of administrative bodies’ legal interpretations) are not subject to parliamentary scrutiny, and since the Awb’s explanation paradigm also does not oblige to justify much to individual decision subjects in terms how underlying knowledges (laws, legal aims and principles) are interpreted and engaged with (nor much else – but that is for later.) This is especially true for the initial explanations, given by initial decision makers. This means that the type of justification that element one promotes is not part of the type of thinking that initial explainers are asked to do, even if the type of social-epistemic engagement that allows them to explain in the first place would be (which it is not). That raises questions about the expectations that the state intends to have with regard to their ‘executive,’ even before the assessment of such expectations’ realism. The sections below run by some additional legal intricacies that are of relevance for element one.

⁶⁶⁸ Section 4.2.3.6

⁶⁶⁹ Section 4.2.3.5

4.4.2.3 *What should reasons represent? The Awb relations of due diligence and motivation*

In calls to improve the humaneness and trustworthiness of the State as expressed in administrative decision practices, the principles of motivation and due diligence are frequently named in tandem. The tandem resembles those that inspired the modeled duties of explanation care: accuracy (/due care/competence), and sincerity (/honesty/trustworthiness.). To recap: the first leg of the tandem saw to the *creation* of knowledge *on* responsible terms; the second to the motivation to *share* knowledge *in* responsible terms. The first leg primarily focuses on methods, the second on dispositions/attitudes of practitioners. The use of ‘primarily’ is key: the point is that the two sides should apply *in tandem*: the one informs the other. For this section’s purposes the required sincerity (what needs to be explained) sets demands for the decisional process that is engaged in; and the methods that are required to engage with need to be reported on to decision subjects.

But in the Awb’s explanation paradigm, the two sides have come ‘undone’ to some extent. In codification history already, it was argued that bad reasons may hide good decisions, and that upon judicial scrutiny, administrative bodies should be allowed to clarify and repair their reasons before the underlying decision is condemned.⁶⁷⁰ Under the Awb’s eventual regime, bad reasons are a (much used) ground for review and appeal, but administrative bodies are indeed not retributed for bad explanation practices in the sense that a new decisional process needs to be started. administrative bodies can repair reasons to the extent that the ground under a decision changes entirely. Together with the rather minimal obligations with regard what needs to be explained in the first place, this reduces the pressure to ‘get it right the first time.’ This arguably reduces the performativity of explanation rules to positively influence the decisional process itself.

This is especially unfortunate in light of the restricted Awb definition of ‘decisions’ that need to be explained at all.⁶⁷¹ The Awb severely limits what actions, behaviors and choices that decision makers engage in, need to be explained, limiting what they are *obliged* to think about in terms of justification. This sets the tandem principles further apart. A typical example of what may be excluded are welfare fraud signal derived from any methods for creating such signals, and investigative practices that are engaged in as a consequence. Another example was the design of a process that checked a person’s eligibility to *start* an official social care eligibility procedure. Both are examples of how salient knowledge making practices are *formally* excluded from the ‘principled’ thinking of explainers.

⁶⁷⁰ Section 4.3.2

⁶⁷¹ Section 4.3.1.1

4.4.2.4 *Reasons on demand: hierarchical relations of decision makers*

The section ends with a consideration about the decision review procedure. The review process became the designated level where decisions are most extensively reasoned. Initial statements of reasons may be kept brief, legal elaboration follows when explainees complain about, or object to, the outcome of their process. This explanation is done by a different decision maker, since internal review functions as a quality and legitimacy check of the initial decision. The check is assigned to decision makers with more legal knowledge and training. This again questions the knowledgeability that initial administrative decision makers are set up to have. An additional issue with this is how first explainers are hereby placed ‘lower in rank.’ They are not the legal expert, and their decision stands to be corrected by those in rank above them. A tentative argument to make here is how this could *dis*-encourage them from intimately engaging with values embedded in law and policy, and with principled justice. They may feel they are not (as) authorized, which may entice them to follow procedure as the safer option. In other words; to not ‘rebel.’ Such conclusions can only be drawn on the basis of further research. But the conclusion can at least be drawn that the Awb’s paradigm does not stimulate their engagement.

4.4.3 Element two: engaging with the social-epistemic positions of explainees

Explainers are obliged to investigate the social-epistemic positions of explainees in relation to the decision-making modalities and underlying (input) knowledge at hand; can explainees be expected to responsibly provide (or have provided) the necessary input, and understand the output? If no (or can’t investigate), rebel.

This element, like element one, obliges to ‘prepare the table’ for the negotiation of the how’s and why’s of decisional outcomes. This time the focus is on how explainees *will be able to* experience a just testimonial process. Explainers need to be able to demonstrate engagement with their explainees social-epistemic situatedness (on individual and group levels) in relation to the larger decisional process and methods: ‘the system.’ This includes engagement with how a system historically treated explainees as a group, and individually. The need to rebel exists when explainers feel their explainees are in no position to participate in the decisional process responsibly.

*

4.4.3.1 *General recognition for element two: the need to look beyond the individual case*

As subject and executive experts, perchance with local experience, national and municipal administrative bodies are assumed to have the right kinds of information positions to make citizens’ particular circumstances and interests ‘count’ in procedures

in a way that legal rules inherently cannot. To do so, civil servants are legally obliged to ‘gather the necessary knowledge about facts and relevant interests’ about the subjects they serve. What counts as necessary and relevant is *principally* unrestricted, and decision makers are end responsible for the quality of individual fact establishment. They are increasingly stimulated to do so from a broad view of the Administrative landscape, and cooperate with other administrative bodies where necessary.⁶⁷²

These obligations can certainly be made to bear on what element two promotes. But persistent legal-systematic and practical obstacles are acknowledged to exist, and case illustrations draw into question to what extent administrative bodies can be expected to take responsibility for the quality of their clients’ information positions. Both things are broadly acknowledged in literature from and about the domain; the quality of citizens’ information positions has had a ‘bad rep’ for many decades. Together with the social and financial power imbalances between administrative bodies and citizens, ‘inequality compensation’ is assumed to be a necessity. The Benefits scandal has foregrounded such concerns, and the further development of several principles (due process, motivation, ‘meaningful contact’) are called for to improve the situation. But both the Administrative ‘ideal’ principles, and the consulted critical literature are mostly geared to the individual’s level. For the Model’s purposes, this is not enough. The sections below discuss some of the mentioned obstacles from this perspective.

4.4.3.2 Information positions of decision subjects: the need to go beyond the ‘complexity’ argument

The element ask explainers to care for the ability of subjects to participate responsibly in decision processes about them. As early domain research pointed out; this means caring for their information positions before, during, and after explanation.⁶⁷³ One obstacle towards this are the acknowledged large legal and practical ‘complexities’ of the domain: not just for explainees but for civil servants and (other) legal experts too. E.g., it is hard for subjects and civil servants to responsibly manage what information about them resides in which administrative bodies’ systems, and make sure it is correct. Information is combined and recombined, mistakes slip in.⁶⁷⁴ The amount and detailed character of rules and procedures doesn’t help, and automation exacerbates all these problems.

But the usefulness of blaming ‘complexities’ is restricted, as is the related tendency to blame ‘legalism’ or ‘formalism’ for the cited problems. Such qualifications either suggest that the system is principally sound but needs procedural redesign, or suffers from an inherent conundrum that can be compensated by allowing administrative employees to engage in more personal and ‘creative’ relations. Both stand in the way

⁶⁷² Section 4.2.3

⁶⁷³ Filet, *Kortsluiting met de bureaucratie : over participatiemogelijkheden van burgers bij het openbaar bestuur*.

⁶⁷⁴ Section 4.2.3.1

of a more profound understanding of what the meaningfulness of subjects' information positions hinges on, and whether the Administrative system is set up to serve them well. For example, the restricted definition of what needs to be explained is arguably unproductive. This was treated above and does not need repeating. An example to add is how the concept and instrumentalization of legal fictions and legal assumptions is challenging: not just with regard to understanding their difference, but because there is no clear regime for what counts as (counter) evidence and relevant interests, and whose burden it is to 'prove' a legally relevant situation. Subjects easily have a hard time understanding how their situation translates legally, and to act in their own interest.⁶⁷⁵

The point was made about policy rules in particular. This salient place for administrative knowledge making is used to reduce motivational burdens. What 'counts' as relevant information and interests tends to be laid down in them, and in absence of clear 'testimonial rules' subjects have no reasons to try and make different information count. In other words, 'accuracy' and 'sincerity' are unrelated again, which arguably does the opposite of stimulating the kind of disposition element two requires.

The phantom vehicles case was a worrying illustration of all these things. Strict and partly hidden policy choices were ambiguously grounded on underlying laws. Apparent impossibilities for subjects to participate responsibly in their processes were willfully ignored; paths to claim disproportionate hardship were actively obscured. Registrees were dependent on the 'personal pity' of a whole range of administrative bodies, judges, the police and the public prosecutor.⁶⁷⁶ And the Social Care Act cases illustrated that when municipal administrative bodies are asked to use their 'trust relations' and local expertise, they do not necessarily design their processes in more accessible ways. Subjects were actively kept from understanding and meaningfully participating in the entire decisional process.⁶⁷⁷

Such cases attest to what appears to be a weak regime in terms of element two. When subjects are practically unable to act in compliance with rules that apply to their situation, or to act in their best interests because they lack the information to do so, the chapter suggested that it can become *principally* unfair to subject them to a processes' rules. But such situations are not what 'disproportionate hardship' clauses are created for. For one, they see to individual circumstances only, and the cases imply group maltreatment. Relief arguably needs to be afforded on the basis of a more principled and related understanding of due process and motivation, and with a clear instruction for decision makers to investigate whether challenges are not incidental but systemic, and which groups in society are hit hardest by them.

⁶⁷⁵ Section 4.2.3.1

⁶⁷⁶ Section 4.2.3.6

⁶⁷⁷ Section 4.2.3.5

4.4.3.3 *What should reasons represent? A case for explainee-insightfulness*

An additional consideration with regard to policy rules and the character of ‘disproportionate hardship’ clauses is of relevance here. The chapter cited how administrative bodies are (since a few years) obliged to establish that there aren’t any special circumstances that would prohibit the application of a policy rule; but they *aren’t* obliged to include this qualification in statements of reasons unless, and until, explainees object to the decision.⁶⁷⁸ Since going through a review process can amount to a significant social, financial, and organizational burden for explainees, this explanation of the explanation rules keeps crucial information away from them.

Which raises questions about the ‘external’ value of statement of reasons. It was discussed how reasons should afford external experts the insight they need to check decisions for compliance with the principle of motivation, due process, and other applicable principles. But in absence of more, and more elaborate codification of the principle, ‘external’ becomes ‘only external,’ which is problematic for various reasons relevant to element two. For one, explainees are excluded from even a theoretical means to know what they have a right to be informed about and explained. Furthermore, decision makers aren’t encouraged to think ‘principally’ with their explainees as theoretical sparring partners—rather with their peers or future adversaries in mind. An additional problem is how this isolates explainees from peer-assistance, too. When reasons repeatedly fail to make decisional choices insightful while declaring that there is nothing of interest to note, explainees’ protracted misfortunes starts to reflect back on themselves, isolating them from support they need and deserve to get.

4.4.3.4 *Making discretionary space work for explainees, with explainees*

The Awb’s regime for investigating and establishing a subject’s facts and interests embodies an anti-oppression functionality in how these activities are subject to relevance to the precise attributed decisional authority, governed by the underlying primary law.⁶⁷⁹ The rule was discussed as (also) problematic with regard to gaining a comprehensive understanding of a subjects’ situation in light of their broader ‘administrative’ lives. This also makes it hard to spot patterns of marginalization across societal spheres. Ameliorating factors exist in a more principled understanding of the Awb regime and decision makers are also explicitly asked to think ‘from’ the perspective of other possibly applicable laws, and from subjects’ individual needs.

But subjects’ own information positions are not named as relevant in either the principled broad space, or the informally as well as formally regulated broad space (e.g., in the Participation law). It is acknowledged that subjects may *not* know which legal eligibility regime most adequately fits their support needs, but it is not

⁶⁷⁸ Section 4.3.3.2

⁶⁷⁹ Section 4.2.3.2

acknowledged that they may have exclusive knowledge of how laws and procedures work out for them or people like them in their situation. In other words, the expectation is framed in a paternalistic fashion. In combination with broad space for decision makers in the preparatory stages of decisions; the lack of a clear regime for evidence; and the unchanged obligation to accept even incorrect facts established by authorized administrative bodies,⁶⁸⁰ this is arguably unproductive in terms of element two.

4.4.4 Element three: practicing interactional justice

Explainers are obliged to practice interactional justice, which entails to recognize explainees as knowers and rights-holders. Explainees should be provided information that is proportionate to their pre-investigated and incidental (self-expressed) needs; their knowledge and understanding of relevant, larger & smaller knowledge making processes at hand should be discussed with them with the aim of promoting their responsible (dis)trust; accessible justificatory sources from outside of the authoritative setting need to be pointed out accompanied by instructions on how to follow up on such leads; explainees need to be afforded information about their rights with regard to the explanation and the decision outcome; the possibility of social pressure needs to be mitigated by e.g. allowing to bring allies or make recordings.

The duties of this element describe the interactional dimension and behaviors that need to be given an explicit place in the testimonial process. If any description goes beyond what a process is seen to need, this will need to be justified in the testimonial record. The inclination of lawmakers to treat much practiced (or ‘bulk’) decisional processes as simple, self-evident, ‘routine’ and predictable has led to sub-optimal explanation practices. The implementation of automation in such cases exacerbates the problems while obscuring their origins.

*

4.4.4.1 General recognition for element three: a momentum for legitimacy?

From early Awb conception onward, legislative, Parliamentary, and scholarly discourse has acknowledged how responsive, honest, and trustworthy interactions with well-informed citizens are key determinants of *legitimate* administrative behavior, and fundamental to the proper functioning of constitutional democracies. The need for accessible human guidance throughout decisional processes, and for ‘reasonably’ rather than just rationally understandable explanations are argued for in critiques of how this ideal is not met in practice, notably in calls to ‘re-humanize’ decisional processes in light of ADM developments.⁶⁸¹ The following sections run by several

⁶⁸⁰ Section 4.2.3.1

⁶⁸¹ Section 4.3.3.3 and e.g., ‘Ongevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen’.

themes to consider the extent that the current explanation paradigm is set up to meet these expectations—and possibly those of element three that aren't named as such. Various subjects were discussed under element one and two for their relevance to the 'preparatory' stages of the modeled duties of explanation care. These discussions will be referred back to while this element adds considerations.

4.4.4.2 Responsive information needs of decision subjects: pointed out, seen, but not captured

The ABAR report that advised the Awb's codification argued how elaborate 'understanding needs' for subjects followed from the duty to provide reasons. Explanations need to serve the cognitive and 'bureaucratic' needs of a decision subject; they need to understandably reveal a decision's underlying rules, the administrative body's explanation and interpretation of these rules, and what these interpretations are grounded on. This includes administrative bodies' interpretations consolidated in policy standards. The report added 'example questions' that administrative bodies can use to prepare their reasons: what goals were aimed for, what interests taken into account, what weights any interests were given, and *why*. Furthermore the report argues that explanations need to be drawn up per individual case, as 'dependent variables' should determine the extent of what needs explaining. An explanation might need to cover any factor, all information, any external advice and, of special interest for element three, *alternative options* and *mutual expectations*. The demands amount to a highly responsive process, much in service of element three. The report however also argued that when such reasons fail, it should not be assumed that the decisional outcome is not properly grounded. With that their advice mainly pertains to the review stage: the place where administrative bodies may repair their reasons. This seems to have been the understanding of the Lawmaker too. But contrary to the advice, parliamentary papers expressed the expectation that the needs and values don't need explicit codification.⁶⁸²

4.4.4.3 Discretionary space for decision makers: a case for formality (revisited)

The preceding sections discussed several cases in which the Lawmakers' professed expectations of responsive, insightful and accessible State interactions did not materialize 'spontaneously.' In part, this could be related back to the primary laws that attributed the discretionary space to administrative bodies. E.g., the Participation Law was built on unrealistic as well as unfair assumptions about citizens, and a strong retributive regime.⁶⁸³ But not just primary laws were to blame. The Lawmaker's positive expectations with regard to the 'beneficence' of the executive were arguably unrealistic, too.

⁶⁸² Section 4.3.2

⁶⁸³ Section 2.3.2

When unrealistic expectations with regard to administrative bodies' policy making and effectuation are a pattern, a lack of State guidance arguably needs to be distrusted. In the phantom vehicles cases and the Benefits Scandal, the State was quick to blame the administrative bodies rather than assume end-responsibility even though they had been made aware of the problems that had arisen. Earlier cited remarks on behalf of the State about experiments with informal review procedures are concerning in this regard as well. Researchers reported instances where explainees' legal aid professionals (if they had them) were actively avoided; researchers were kept away from informal contact moments, and case files did not include explainers' notes of what happened in them. The State acceded that this rendered the processes unaccounted for, and yet that there was no reason to assume 'worst case scenario's' (in which explainers exercise undue pressure to accept a decision) would unfold. In an Ombudsman study on informal *complaints* procedures, where interaction (and so, social pressure) itself is the subject, similar concerns of legitimacy were raised: the absence of written confirmations of commitments and agreements, and of properly reasoned decisions.⁶⁸⁴

These findings raise questions about the affordance of the codified explanation paradigm, and what this should mean for the further development of the principles of motivation and due process. Not just to promote responsible State behavior, but to strengthen the institutional expression of the right kind of norms more broadly, and educate 'ourselves' as society. From that viewpoint, one may question the purpose (and usefulness) of the development of a new principle of meaningful contact with the government. The principles of due diligence and motivation together already express most (if not all) of what element 3 promotes, still various 'proper interaction' instructions in the Awb paradigm are badly complied with. Creating another principle arguably downplays the other principles' meaning, and pushes the right kind of treatment into the realm of informality and discretion even more.

An argument was cited in favor of a 'duty of care' over a new principle, since a duty of care at least obliges administrative bodies to book results.⁶⁸⁵ This section would agree, but in light of the persistent problems also argues to bind the duties legally to the Awb's explanation paradigm. This would add rules in what the thesis, with the author, recognizes as an already highly complex legal landscape. But as was discussed under earlier elements, 'complexity' should not be used as an excuse for reducing or omitting necessary instruction.

⁶⁸⁴ Section 4.3.3.5.

⁶⁸⁵ Section 4.2.2.2, Claessens, 'Het (on)nut van een recht op toegang tot de overheid als nieuw algemeen beginsel van behoorlijk bestuur'.

4.4.4.4 *What should reasons represent?*

The weak relations of due process and motivation in the Awb's explanation codifications are unproductive with regard to element three's promoted values, which embody how explanation is relational and needs to be a process. In light of subjects' purportedly low 'literacy' of Administrative law and policy, the recognition of explainees as 'knowers' and 'rights holders' would benefit from an explanation of their (broader) due process and explanation rights. The obligation to provide reasons becomes a poor understanding of what a testimonial process is when explanations are 'just' expected to legitimize an outcome, and not to testify to how an Administrative body has been an understandable and trustworthy decision partner.

The lack of retribution for administrative bodies who give 'bad reasons,' discussed before, is problematic in this light as well. For one, in terms of promoting the right kind of trust and distrust. Secondly in light of how the same 'corrective' space is not afforded to explainees, themselves, which exacerbates the power imbalance between them. The fact that administrative bodies are obliged to use information that they know, or can know, to be wrong is an illustration of both these things, and asks civil servants to 'reason away' important circumstances that deserve to be considered.⁶⁸⁶ Put differently, such obligations cannot possibly produce reasoned statements of how unjust knowledge making practices are avoided.

4.4.4.5 *Reasons on demand in relation to the principle of motivation's 'compensation' function*

Several aspects of the Awb paradigm work out reductively with regard to the reasoning of initial decisions. Explanations can be legally withheld in case of (typically) positive eligibility decisions, and provided on demand in such cases. For positive and negative decisions alike, research shows that the initial statements of reasons that *are* given tend to be quite minimal: keeping to the minimal demands of codification, rather than the spirit of the principle of motivation. Furthermore, policy rules are used to lessen motivational burdens and the review procedure is used as an additional 'reasons on demand' clause. With this, the review procedure became the main place where elaborate reasons are *made*, as well as the main place for quality control of administrative decisions.⁶⁸⁷ This focus on the review stage is hard to understand in light of the principle of motivation's aim of reducing power and information inequality. The lens of element three helps to see how. Element three promotes a process that *always* aims to improve explainee information positions, which requires meaningful explainer-explainee exchanges. It also obliges administrative bodies to address social dynamics in what is a highly unbalanced power relationship. Asking subjects to take burdensome action to get access to explanation is unhelpful of itself, but becomes more so when

⁶⁸⁶ Section 4.2.3.1, the system of 'single authority' information establishment.

⁶⁸⁷ Sections 4.3.1.2 and 4.3.3.2.

explainees don't have sufficient information to ground a decision on whether to file for review. Section 4.3.1.2 above already problematized the 'soft' legal response that administrative bodies get upon providing poor quality reasons that cannot ground the decision, to add here is that it doesn't respect explainees as rights holders. That section also emphasized how the quality of reasoning should be made to matter in a legal paradigm that relies on language to assess, but also challenge legitimacy. Language is the tool with which to contest the decisional process that a set of reasons represents. For this, reasons needs to give sufficient insight. If the argument is of bad quality, this should have consequences, same as arguments *against* a contested decision are only accepted on the basis of their quality. Downplaying the value of reasoning over actionable conclusions also opens the door to arguments of ADM times that principally downplay the usefulness of causal understanding.

But the *type* of decision maker that explainees interact with in review matters too. These are 'the lawyers of the house,' and depending on the kind of procedure modality that is chosen, individual explainees may be faced by a whole committee of them. In other words, the character of the process changes in terms of how 'explanatory knowledge' is made, but also in terms of power (im)balance. Seen from the aim of element three, the elements' suggestion to 'bring peers' or make recordings is not enough; expert support should be provided. At the very least, records about the process need to be drafted, so that peers as well as experts can assess the review process's quality.

4.4.5 Element four: creating records

Explainers need to create records of explanation practices. These should be understood as truthful accounts of the testimonial exchange as it was prescribed under element three. Therewith the record should express how all previously described duties were attended to, or provide reasons for when they were not. The records need to be shared with explainees, and made available for outside scrutiny in accordance with rules that govern the decisional domain and relevant privacy and data protection regulation.

These record-related duties are meant to produce more comprehensive accounts than the 'statements of reasons' that are typically the outcome of decisional processes. This acknowledges how explanation is a knowledge making practice itself, and therewith a place or conduit of possible oppression. Comprehensive records can sustain progressive development of decision and explanation practices across time and domains.

*

The element's description already implicates how what it envisions goes beyond typical legal statements of reasons. It does not say it should replace them; but that additional things deserve to be recorded. The wisdom of such a division can be questioned for this domain however. For one, the content and required quality of (initial) statements

of reasons in this domain is lacking. For the decision subject, the records that they are currently provided are not sufficiently usable: not to discuss their situation with peers or their legal advisers, not to support a decision to file for review – a decision they need to make on their own after legal aid for review procedures was terminated. The fact that more elaborate explanation processes are located at a level that only unlocks after explainees start an additional process is a second consideration. The fact that there are no guarantees that proper reasons are put to record *in* review is a third, especially in light of how the State downplayed concerns about the legitimacy of informal review experiments. All these things suggest that responsible explanation processes *per se*, and as a rule, are not taken seriously enough in the current governance paradigm. For external experts, the general public, and researchers on Administrative explanation practices, locating element four's demands in additional records is still useful, but the point of the Model is to serve the explainer-explainee relationship directly, and not (just) indirectly.

4.5 Chapter 4 in a nutshell

This chapter performed an analysis of the basic legal explanation paradigm of the Administrative domain, in terms of the modeled duties of explanation care. Over different sources, the study revealed substantive recognition for many of the values and aims embedded in the Model's obligations—saliently in pre-explanation codification discussions, and in ongoing descriptions of how the domain needs to do better to meet such expectations.

Very long-standing concerns exist about the lack of insightful, trustworthy and understandable relations with decision subjects, which raises questions about consecutive governments' appetite for progress. The Benefits Scandal added concerns about administrative bodies' wrongful interpretation and application of underlying laws. Some critiques engaged with the in/justice potential *of* these underlying laws, in line with earlier acknowledged political 'mishaps' in for example welfare laws. The character of Administrative judicial scrutiny, finally, is under renewed scrutiny itself for how it fails to deliver justice when it needs to.

The domain research surfaced various relations of these concerns with the Awb's explanation paradigm, revealing weak points that arguably require attention. Chapter 6 brings the chapter's findings to ADM explanation discussions, calling attention to how such weak points, left unattended, stand to weaken 'above ground' reparations in AI-informed times.

This 'nutshell' does not summarize all the chapters findings, but points out the main themes in which problematic aspects of the explanation governance surfaced.

Lack of instruction for the principled use of attributed powers

In their executive and legislative capacities, administrative bodies are expected to make citizens' actual and contextual needs and situations count in ways that general laws cannot. They need to make sure that rule applications don't lead to disproportionate hardship, and are relied on as a stable expert force against (possibly oppressive) political 'whims.' The chapter research surfaced various concerns with regard to the principles and values that administrative bodies are (not) instructed to engage with in the use of these powers. Illustrations of how they 'made the worst' of laws that were the product of ever harsher, ever more discriminatory political climates are non-incidental. Policy design, away from parliamentary scrutiny, is also used to *restrict* what subject's facts and interests can be made to count, and to (severely) reduce burdens of explanation and justification. When administrative bodies are condemned for such practices in Administrative appeal procedures, they appealed to Administrative supreme courts to fight for their ways (and tend to win.)

Still in calls for more 'humane' treatment of (especially) citizens in need of assistance or support, there is a tendency to blame 'formality,' legal administrative complexities, and (digital and analog) bureaucratic system intricacies. 'Informal' discretionary space is seen to be lacking. But on the basis of fundamental legal principles, especially those of motivation and due process, administrative bodies already have the space to 'do justice.' *Formal* instructions, i.e. further codification to use this space and use it rightly are arguably what are missing. Important for such instructions is to understand what 'disproportionate hardship' looks like on non-individual levels, and to engage with underlying laws at hand for how they produce it. This arguably takes education, and work force diversification, in light of the denials of discriminatory wrong-doing on Dutch institutional levels.⁶⁸⁸

Legal intricacies not conducive to meaningful subject participation

Several legal intricacies reduce the quality of explainees' information positions in what is acknowledged to be an already highly challenging environment to understand. For example, the Awb severely limits what steps of decision processes need to be explained to them. 'Investigative' stages of decision making, prominent places for wrongful group treatment, are typically excluded. This legal design also limits what civil servants are obliged to think about in terms of explanation, which is precisely contrary from the Model's point of view.

This situation is not ameliorated by other rules. The Awb lacks a bespoke codified regime for what counts as evidence, and for the distribution of burdens between State and citizen with regard to making sure the right things are considered. In such a space,

⁶⁸⁸ Brenninkmeijer, 'Welke lessen zijn te trekken uit de kinderopvangoeslagaffaire en de problemen bij uitvoeringsorganisaties?'

it would make sense to make sure that decision makers are instructed to explain to subjects what their *principled* due process and explanation rights are, since these are acknowledged to go beyond what can be gleaned from the codified rules. This bridges to the next theme.

Making reasons meaningful for explainees: relating due process and motivation

In calls to improve administrative practices, the principles of motivation and due diligence are much named in tandem. But their codification is mostly separate. Per the Awb, explanations are set up to state the legal grounds for an outcome, not to testify to how an administrative body has been a responsible and trustworthy decision partner. As stated, a principled understanding of ‘motivation’ would require that much more is justified. Statements of reasons’ ‘external value’ is legally explained as a requirement that lets judges and other ‘experts’ check an administrative body’s compliance with applicable legal principles. The fact that typical, initial statements of reasons, the ones that go out to explainees, don’t allow for this at all is not considered as problematic. This leaves explainees with statements that don’t testify to their State interactions; that ‘reason away’ important clues about the quality and justness of their processes.

At the same time, a *new* principle is in the making after the Benefits Scandal revelations: that of ‘proper State relations.’ That mission is hard to understand when the principles of due diligence and motivation together already cover most, if not all of the grievances—and would ‘just’ need more codification.

The burden on and of review

Administrative review is the level where decisions are most extensively reasoned. Administrative bodies can repair their reasons to the extent that the grounds under an unchanged decision changes entirely. Several problematic aspects of this were engaged with. Among them: this system reduces the pressure to ‘get it right the first time,’ reducing the power of explanation regulation to stimulate proper decision making. Making a properly reasoned decision dependent on filing for review also amounts to an unfair burden on explainees, especially with not much initial decisional information to go on. In addition, ‘first explainers’ are placed lower in rank since the review team is more legally trained. This could *dis*-encourage these initial decision makers to engage in less clearly instructed, principled engagement with law and policy. More fundamentally, it was considered that the quality of reasoning should be made to matter in a practice that relies on language for its expression of justice.

Interestingly, in review procedures as well, ‘informal’ discretionary space was shown to be used in useless ways, raising questions of power abuse and legitimacy. The same was true for ‘informal’ complaints procedures. Authors were cited to argue for

the recording of explanation conversations as standard policy, and for obligatory written confirmations and properly (re-)reasoned decisions.

Care to explain?

5 Meaningful information positioning and legal explanation rules for General (medical) Practice

5.1 Introduction

5.1.1 Function and value of the General (medical) Practice domain study

The development and implementation of data science and AI for and in the medical domain is fast-paced and looked to with expectations and concerns,⁶⁸⁹ not least with regard to explainability.⁶⁹⁰ To reiterate what was mentioned in the methods section of the thesis's introduction, all the explanation-related concerns that were discussed in the introductory chapters are prominent in discussions on medical decision making: from individual (patient) rights to responsible participation, from explainers' own understanding to the value of (especially causal) explanation as a concept. Decision subjects in this domain are (generally) in vulnerable positions as a given, and the information balances are already large. But perhaps more so than in other decisional domains, questions with regard to the inscrutability of AI-infused knowledge feed into discussions that were already going on about the value, function, and possibility of explaining medical knowledge to patients—and of doctors' own understanding.⁶⁹¹ The regulation of explanation is also relatively young compared to that of the Administrative domain for example, and the role of law less settled.⁶⁹²

It is therefore interesting that the domain's fundamental principles are much cited as useful informers on how to proceed in AI governance, also with regard to 'transparency.' This was one reason to select the domain and investigate it with the aim to answer its part of the thesis's third research question: "how do existing legal rules of two seminal regulated explanation domains promote responsible (non-oppressive, information position improving) explainer behavior?"

689 Robert Sparrow and Joshua Hatherley, 'The promise and perils of AI in medicine' 17 (1 December 2019): 79–109, <https://doi.org/10.24112/ijccpm.171678>; Tamar Sharon, 'When Digital Health Meets Digital Capitalism, How Many Common Goods Are at Stake?', *Big Data & Society* 5, nr. 2 (1 July 2018); Powles and Hodson, 'Google DeepMind and Healthcare in an Age of Algorithms'; 'Algorithmic Impact Assessment: A Case Study in Healthcare' (Ada Lovelace Institute, 8 February 2022), 29–30.

690 Rune Nyrup and Diana Robinson, 'Explanatory Pragmatism: A Context-Sensitive Framework for Explainable Medical AI', *Ethics and Information Technology* 24, nr. 1 (2022): 13.

691 Katz, *The Silent World of Doctor and Patient*. 1984.

692 As will be discussed in detail in this chapter.

The value of this chapter's analysis is however not dependent on whether AI is embraced or stalled in the domain. This rather lies in the domain's foundational standing and the vulnerability of patients. In the doctor-patient relationship, the information inequality gains weight through the social dependency of explainees for their well-being and thriving. The scoping choice for General Practice (GP, or Family Medicine as it is also called overseas) was made because, like the Administrative domain, this practice serves all citizens: the doctor-patient explanation relations is a commonly experienced one. GPs in The Netherlands are gatekeepers of more specialized practices and patients need their referrals to proceed.

The domain is a prototypical expertise-based domain, which adds value for Chapter 6 that draws lessons from both domains, from the investigation of two useful 'prototypes,' together. The field is also known for how wrongful knowledge practices are historically rife within it. Examples from the larger medical domain populate epistemic injustice literature *en masse*. Legal philosopher Marcum discusses the need for conscientious practice in this light, discussing discriminatory biases in all of medical knowledge making as a particularly widespread handicap whose "deleterious impact" may result in "cognitive myopia", disabling a doctors ability for accurate diagnosis.⁶⁹³ The awareness of these issues functions as a backdrop for the chapter itself which does not (again) seek to provide evidence that these issues exist. That said, the politics of medical knowledge in general are of course well within focus.

The chapter proceeds as follows: the first part of the chapter (section 5.2) offers a functional characterization of the GP domain as a field of knowledge-and-decision making, including a characterization of GPs as knowledge and decision makers. Like the previous domain chapter, it categorizes the findings in three sub sections ('what, who, how'). Together, these provide a basic understanding of GP's social-epistemic positions, and the attributed, as well as self-built role authority of GPs *vis-à-vis* Dutch patients. The second part of the chapter (section 5.3) discusses the field's basic legal explanation obligations, and places them in the larger governance context. Norms around explanation are established through public and self-regulation: law, medical and professional ethics, professional standard setting. The chapter embeds the legal standard setting by providing illustrations of how the other fields inform, resist, explain or adopt these rules. The third part (section 5.4) brings these two parts together, relating them through the epistemic in/justice informed lens of the modeled duties of explanation care.

5.1.2 Research and reporting choices

This section reiterates some research choices with regard to the GP domain that were discussed in the methods section of the thesis's Introduction, and adds some reporting details. The main important thing to (re-)mention with regard to the investigation

⁶⁹³ Marcum, 'Clinical Decision-Making, Gender Bias, Virtue Epistemology, and Quality Healthcare'.

presented this chapter is that the absence of case descriptions (either in case law or other literature) may leave the reader feel they need more actual practice insight. The choice follows, firstly, from the role of law in this field: although law has been a necessary instrument with regard to explanation regulation so far, the legal rules leave most detailed instructions up to the medical and ethical norm setting dimensions of the field. The relation between law and the other governance modalities was therefore allowed to take up space. In addition, the historical paradigm shift from ‘doctor knows best’ to ‘informed consent’ gets ample attention, and is embedded in an elaborate treatment of what GP practice and (their) medical decision making entails. This approach affords insight into what are still developing rationales with regard to explanation in the field: it includes important differences of opinion about what *can* and *should* be explained and what the role of law should be in guiding and prescribing this. For the thesis’s purposes, this was considered to be sufficient.

The descriptive ‘what, who, how’ parts of this chapter are therefore larger and more detailed than their counterparts in the Administrative domain chapter. There are several additional reasons for this. For one, unlike the Administrative domain, the GP domain did not get a provisional introduction and the ‘red thread’ of the Benefits Scandal is not complemented by a GP scandal or a more general medical ‘case.’⁶⁹⁴ In other words, there is no additional embedding of the GP domain in the thesis and everything that needs to be discussed is discussed here (although, with reference to the ‘backdrop’ remark above, medical injustice did already receive attention.) Domain-intrinsic reasons exist in the character of this domain, and of its decision makers/explanation providers.

To start with the first, medical practice is ‘all about’ knowledge making and vast stretches of the landscape are cultivated by (who gets to contribute as) experts. No democratic, constitutional negotiations ground the produce of medical knowledge practices. Unlike the administrative domain, medical knowledge is not considered to be ‘known’ let alone principally understood by non-experts, as for example expresses in much heard arguments about how people ‘trust’ rather than understand, their GPs. To go beyond this simple presentation and arrive at a sufficiently adequate understanding of the domain, the relation between knowledge, and what has become *shared* decision making needs a more thorough discussion. The choice was made to focus on general-medical consultations rather than specific medical or legal states, such as end-of-life decisions, minors, or persons with incapacitating mental states. In addition, in explaining what GP medical decisions are, more attention went out to the diagnostic phase of medical decision making. This is the dominant type of decision making in the domain, and a useful ‘vehicle’ for making the sociality and politics of medical knowledge as it touches everyone, insightful.

⁶⁹⁴ And as was described in the introduction to the thesis, the Administrative domain was done first, and *inspired* to go for a more fundamental ‘re-idealization’ of legal explanation. The Administrative domain research was somewhat ‘cut short’ because of this.

Secondly, the *person* of the GP is discussed more in-depth. Unlike (ideals of) ‘impersonal, interchangeable’ bureaucrats, GPs are expected to be invested personally. They are legally and ethically, individually and morally responsible, and a thicker description of this weighty social-epistemic position was warranted in order to understand the ambition of the domain’s legal explanation rules in light of it.

Lastly, following from what was discussed above with regard to the extent of the chapters’ report on what GP’s eventually need to explain: the modeled duties of explanation care-analysis in this chapter takes a more ‘modest’ position *about* what happens in practice, at least compared to the Administrative domain. An additional reason is that the domains’ record related obligations were not studied, and in any case are not made accessible because of privacy and data protection law obligations. The analysis however provides a thorough basis for further empirical research, which could usefully inform the implementation of the Model in the legal governance structure of the domain.

Some notes on literature

Literature about the field of study, and on subjects and objects within it were collected from a range of research fields: medicine, law, (medical and professional) ethics, (other) philosophy, social medicine. The materials included books, handbooks, journal articles, institutional reports and education materials. Among the books are two large longitudinal studies on different aspects of Dutch GP practice. Both books include a thorough literature analysis of two different major sources: *Huisarts en Wetenschap* (‘GP & Science’) that in its early years was much concerned with defining, and promoting GP as a specialist profession; the other *Medisch Contact* (‘Medical Contact’) which was the more practice-oriented magazine. One field that was not sourced was medical communication (methods). Although this rich literature can obviously teach a lot about explanation in the domain, the quest is for *what* needs to be explained rather than how to do this. *The Silent World of Doctor and Patient* (Jay Katz 1984) stands out in the list of sources because it is from the US. A seminal and much cited work on explanation in medicine, it describes and argues the shift from the Doctor Knows Best to the Informed Consent paradigm, specifically with the goal of shared decision making which has by now become the norm also in The Netherlands. It combines medical, ethical, and legal knowledges and their histories, and its relevance is not contained to the US context.

The scope of legal explanation rule research was delineated to the main explanation rules of the Medical Treatment Agreement Act (‘WGBO’). The analysis was kept to (this) Dutch Law. There are additional explanation duties in other applicable laws, such as in the 2016 Quality, Complaints, and Conflict in Care Law, article 10, under 2: “The care provider informs the patient about the existence of scientific evidence about a treatment’s efficacy.” These and other laws were not studied. Although they

apply to medical decision makers, they are not where the core explanation duties and rights are established and not discussed as such in legal handbooks. That said, valuable instruction may be in them, and any other, or further, effort to further the domain's explanation paradigm would certainly benefit from drawing all applicable laws together.

For the chapter's (and the thesis's) purposes it was furthermore not considered necessary to place the WGBO in its international (treaty) context, especially since – like in the Administrative domain chapter – case law in which the international context is primarily made to bear was not a major source. Most important explanation case law was not about GP practice, and the WGBO had recently gone through a major revision, incorporating important case law developments. The handbook of Dutch Health Law followed suit, and both the WGBO's explanation rules and the Handbook were studied for salient developments. The WGBO's update was informed by several background studies, undertaken by different research bodies. These were included as research materials. Added to these core sources were a collection of inaugural lectures of Dutch health law professors 1971-2011. The collection looks back on developments and forward to the future of Dutch health law, and as such helped to characterize it.

5.2 General Practice decision making in The Netherlands: a functional characterization

5.2.1 Elementary GP decision making: diagnosis ('the what')

5.2.1.1 Introduction

This first section discusses 'what' typical GP decisions are, or rather, *a* what. Priority is given to diagnostic conclusions as the typical outcome of GP consultations. The section could also have selected the concept, or paradigm of Shared Decision Making (SDM) as a prototypical 'what' in the sense of what happens, but chose to treat that in the 'how' category. The reason is that SDM practice is a relatively young method of coming to medical decisions.⁶⁹⁵ Understanding SDM benefits from a preliminary discussion of what medical knowledge making comprises and how medical expertise, typically described as the doctor's or 'medical technical' side of SDM practice is social and political too. It also benefits from the characterization of General Practitioners as medical knowledge makers that is provided in the middle of the three 'what, who, how' sections. The choice therewith supports a functional buildup of GP practice as a social-epistemic enterprise.

⁶⁹⁵ And as will be discussed, it was not naturally endorsed by professionals. In the words of an opponent to SDM's assumed ideals, SDM is a 'young feel-good term' that misses the point of medical expert responsibility, comparable to airline pilots who would share decisions about route and speed with passengers based on their preferences. 'Shared decision making is drijfzand', *Medisch Contact* (blog), 9 November 2016.

Diagnosis is the typical starting point of a presented or presenting patient's medical health 'incident.' The chapter understands the diagnostic phase as the colloquial and / or physical investigation into *what* ails a person, finding out *why* this is so, and *whether and how* a person can benefit from further diagnostics and/or treatment, including medication. These activities are not necessarily cumulative: especially in the GP domain, why's frequently remain unresolved. Many of the most common ailments (such as headaches, abdominal pains) are badly understood, and as will be discussed, Dutch GPs are known to operate on the premise that the body resolves many of them without further action. This also features in the hesitancy of Dutch GPs to provide the required referrals to specialists for further diagnostics and treatment, at least compared to their peers in neighboring countries. And so, a first GP consultation typically concludes with the recommendation to return if the complaints don't resolve within two weeks.⁶⁹⁶

The gatekeeper function with regard to specialist care already means that Dutch citizens' health care experiences are influenced by GP diagnostic practices to a considerable extent. Authors do call attention to the fact that the diagnostics that Dutch GP's *can* perform and pursue are significantly predetermined in other fields. E.g., national disease screening programs and basic insurance coverage of treatments and diagnostics are made in Public Health policy; decisions about what ailments to research and develop treatment for are in public, academic and commercial hands. But as will be discussed, GPs are also relied on to *inform* public health policy and argue for diagnostic (and other health) needs of their patients that aren't met by it, and to remain critical consumers of (commercial and other) diagnostic and treatment innovations.⁶⁹⁷ With this, the 'sociality' of GP diagnostics is introduced.

5.2.1.2 *The sociality of diagnosis*

Medical knowledge that GPs make use of in their diagnostic practices is made in different ways, by different actors, who interact with myriad technologies. These practices serve up different types of insights, leading to different understandings.⁶⁹⁸

696 For example, Johan Legemaate, 'Nieuwe Verhoudingen in de Spreekkamer: Juridische aspecten', *Achtergrondstudie RVZ-advies*, 7 February 2013, 7 under 12.; See also Annemarie Mol en Peter van Lieshout, *Ziek is het woord niet: medicalisering, normalisering en de veranderende taal van huisartsengeneeskunde en geestelijke gezondheidszorg, 1945-1985.*, 2008 This doesn't mean that nothing else of importance happens in GP consultations, of course: a large array of small treatments are administered, too, on the premises or during house calls: from wound care to placing coils, from administering curative medication to the assistance of patients who end their lives.

697 In the words of the Dutch 2020 Medical Training Framework that informs all national medical training programs, "a doctor should not only be a medical expert, s/he has to be a communicator, a collaborator, a leader, a health advocate who acts in society's interests, a scholar who thinks in scientific and moral-ethical terms, and a professional who shares knowledge, attitude and skills with others." 'Raamplan Medical Training Framework' (Nederlandse Federatie van Universitair Medische Centra (NFU), May 2020), https://www.nfu.nl/img/pdf/20.1577_Raamplan_Medical_Training_Framework_2020_-_May_2020.pdf.

698 Marc Berg and Annemarie Mol, red., *Differences in Medicine: Unraveling Practices, Techniques, and Bodies*, Body, Commodity, Text (Duke University Press, 1998), 6.

Since GPs (or any doctor) can't acquire a definite or complete understanding of what ails their patient,⁶⁹⁹ they need to make choices with regard to what knowledge to make use of, how to appraise it, and how far to take their investigations in order to make such choices. Concerned with the sociality of medical knowledge practices, Rosenberg argued that categorizing such choice making as the medical-technical side of consultations attributes medicine with more 'certainty' than it deserves, and sketches a simplified picture of consultation: as a place where diagnosis starts and stops.⁷⁰⁰

A more comprehensive description of diagnostic consultations would reveal how similar symptoms can lead to different diagnoses. This happens for various reasons even before the 'patient side' of consultations, i.e. experience, needs and preferences are taken into account. In *The Body Multiple*, Mol shows how physical phenomena (indeed) mean different things for different sufferers, but also elicit more than one interpretation, or diagnosis, from medical experts dependent on the expertise of the physician that looks at the problem.⁷⁰¹ And so, it matters who GPs choose to refer their patients to. Smith's 1981 social study of 'black lung' disease illustrates how the choices and engagement of personal physicians matter.⁷⁰² It recounts how disabled coal miners were subjected to their company's preferred diagnostic track to 'score' their affliction, and determine their financial compensation. The company had selected X-rays over another available method, Respiratory Disability testing, although (or because) X-rays could not reveal the extent of experienced disability that Respiratory testing could. Company-independent GPs supported miners who protested against low, X-ray-derived scores, and brought their expertise to bear in court.⁷⁰³

699 As medical philosopher Wartofsky argued: "[m]edicine .. constitutes one of the basic and earliest forms of human knowledge, sharing with the other forms certain features and constraints having to do with human learning and the development of skills." Marx W. Wartofsky, 'What Can the Epistemologists Learn from the Endocrinologists? Or Is the Philosophy of Medicine Based on a Mistake?', in *Philosophy of Medicine and Bioethics: A Twenty-Year Retrospective and Critical Appraisal*, edited by Ronald A. Carson and Chester R. Burns, Philosophy and Medicine (Dordrecht: Springer Netherlands, 1997), 55–68.

700 Charles E Rosenberg, 'The Tyranny of Diagnosis: Specific Entities and Individual Experience', *The Milbank Quarterly* 80, nr. 2 (June 2002): 256.

701 Annemarie Mol, *The Body Multiple: Ontology in Medical Practice* (Duke University Press, 2003); and the affordances, traditions, cultures and organizational aspects that belong to their trade. See also Van der Laan and Olthuis on Alzheimers' disease. Anna Laura van der Laan and Gert Olthuis, 'Speuren, puzzelen of afstemmen. Alzheimerdiagnostiek', in *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen (Boom Filosofie, 2013).

702 Barbara Ellen Smith, 'Black Lung: The Social Production of Disease', *International Journal of Health Services* 11, nr. 3 (1 July 1981): 343–59; The study is cited by Berg and Mol, who wonder if the "overly activist tone" of the published article is to blame for its low take-up afterwards. Berg en Mol, *Differences in Medicine: Unraveling Practices, Techniques, and Bodies*, 2.

703 Mol and Berg cite the example to show how social struggle "continues right into the 'heart' of biomedicine" rather than that it adds a dimension to otherwise technical, socially neutral knowledge. Smith, 'Black Lung'; See also Rosenberg who points to how (other) problematic disease entities were justified in terms of their "material mechanism," too. Putative ailments, he calls them: railroad spine, soldier's heart, shell shock.. Rosenberg, 'The Tyranny of Diagnosis', 246.

Mol also argued that different diagnoses amount to different *judgments* about patients' lives: a lab method may reveal a patient as genetically burdened (tough luck), a preventative diagnostic track may emphasize the role of un/healthy behavior (raising questions of blame).⁷⁰⁴ Rosenberg's critical discussions of historical diagnostic developments presents diagnostic labels as "ideas about disease" in the guise of descriptions of biological states.⁷⁰⁵ Both notions are important in light of how diagnostic concepts (and labels) are used as 'organizing principles' in the societies that patients are a part of, and feature in patients' as well as societies' self-understanding.⁷⁰⁶ Since medical knowledge is in perpetual development, the consequences of 'diagnostic progress' can improve, but also disrupt patients' social-economic positions.⁷⁰⁷ This is especially pertinent in societies where much resolution is made dependent on being able to present a medical state.⁷⁰⁸ A Dutch example exists in Chronic Fatigue Syndrome (CFS.) Between 2005 and 2008, the status of 'disease' for CFS was established by Dutch Parliament, negated in Public Policy on disability pensions; acknowledged in judicial procedures, negated as 'consequential diagnosis' by the Public Health Secretary.⁷⁰⁹

Dutch professor Dehue studied various highly GP-relevant 'diagnostic cases.' Among them Attention Deficit Hyperactive Disorder (ADHD) which is established on the basis of a simple psychological test. GPs are allowed to prescribe amphetamine-like medication (e.g. Ritalin) to diagnosed persons—saliently children. In line with Rosenberg's arguments, Dehue is concerned that inconclusive judgments about a set of behavioral symptoms got to 'reify' through public and medical discourse. The ADHD

704 Mol, *The Body Multiple*, 180 She also emphasizes how these different enactments still 'hang together': they are interdependent, which is why she chooses the term multiple over plural. Different experts still 'share' a patient. And may share available budgets, facilities, technologies, cultures, and biases. They are also still dependent on each other's findings and experience.

705 See also Glas, who speaks of diagnoses descriptive, legitimizing and explaining/validating functions Gerrit Glas, 'Ziekte en stoornis in de psychiatrie', in *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, and Gerben Meynen (Boom Filosofie, 2013), 138; Rosenberg cites the well known and infamous struggles around 'homosexuality' as a mental disorder in the 70's and 80's editions of the American Psychiatrists Association Diagnostic and Statistical Manual. The book is better known as the 'DSM' - which, in its 5th iteration, is used by Dutch mental health professionals, Dutch insurers, and in Dutch courtrooms to determine a persons' status, eligibility for treatment insurance, and blameworthiness. Rosenberg, 'The Tyranny of Diagnosis'; See also Maartje Schermer, 'Wat is Ziek, Wat is Gezond?', in *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen (Boom Filosofie, 2013).

706 Rosenberg, 'The Tyranny of Diagnosis', 255.

707 Jules Montague, 'What Happens When Doctors Change Your Diagnosis?', *The Guardian*, 11 June 2018, sec. Life and style, <http://www.theguardian.com/lifeandstyle/2018/jun/11/what-happens-when-doctors-change-your-diagnosis>.

708 Trudy Dehue, 'Definities die oorzaken worden', in *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen (Boom Filosofie, 2013), 184.

709 Schermer also discusses the category of substance addiction, a problem that received all kinds of different labels in different times and places: bad or weak character, criminal behaviour, a neurological disease. Schermer, 'Wat is Ziek, Wat is Gezond?'

symptoms, themselves judgments about what proper functioning is, *became* diseases that *cause* behavior.⁷¹⁰

The prescribed medication toned down children's' acting out, at home as well as in the classroom. Fast growing numbers of parents and teachers presented children for ADHD diagnoses. Eventually social, and even financial incentives to 'get diagnosed' established for young persons themselves.⁷¹¹ They could get 'extra time' during exams; and when pupils' non-diagnosed peers discovered that the medication improved their concentration too, children started selling their pills—a problem that proliferated quickly.⁷¹² Dehue warns for the obfuscating effects of 'reifications' such as these. When a society labels groups of persons' states, or traits, as problematic, when they are seen to hinder societal functioning ('depression' is another one of her examples) and are medicalized, *societal* causes risk to be ignored, and societal resolutions unexplored.⁷¹³

To conclude: the studied literature describes diagnosis as an outcome of individual medical consultations—but also as an outcome of discussions and debates on the terms and definitions of normative concepts such as sickness, well-being, and what is 'normal' functioning. These debates are held on many levels, and there is no agreed distribution of powers between them. In 1948, the WHO defined 'health' as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." Although welcomed as a progressive move at the time,⁷¹⁴ the definition later became criticized for its use of the overambitious 'complete', and how that brought in the risk every possible type of non-well-being would be medicalized. Alternative conceptualizations of health in terms of "adjustment, resilience, and

710 And in later stages get to be discussed as problems that reveal themselves in different ways in different people, at which point the diagnosis diversifies in sub-types. Dehue, 'Definities die oorzaken worden'.

711 Dehue.

712 Dehue; Eventually, the market for non-prescription varieties of 'brain pills' and their counterparts, 'sleep pills' was boosted too, see for example such as [braincaps.nl](https://www.braincaps.nl). These are pills whose workings and hazards aren't clinically tested. They too became popular quickly, and were even sold in university vending machines for a while in between "rulers, condoms, and antacids." The citation is of an elderly pharmacist who spoke up about the moral hazards of such choices. Among other things, he argued that a University-issued suggestion that the complex, multi-dimensional phenomenon of 'concentration' is captured by a para-pharmaceutical pill encourages inevitably disappointed students to go and seek 'the real thing' from doctors and pharmacies, therewith feeding into the already dangerously high ADHD-medication demand. Marc Kruyswijk, 'UvA stopt met verkoop concentratiepillen na klacht', *Het Parool*, 6 September 2018, sec. Voorpagina, <https://www.parool.nl/gs-bfaeef85>.

713 Dehue, 'Definities die oorzaken worden'.

714 Another WHO example pertains to osteoporosis, problems due to the decline of calcium in older people's bones. Long seen as a normal effect of aging, as of 1994, it is a WHO acknowledged disease, reflecting a different view of what old age should mean.

maintenance,” were proposed⁷¹⁵ and a group of Dutch doctors defined health as ‘the ability to adapt and self-manage.’⁷¹⁶ These notions are important (and will be returned to) with regard to what GPs’ explanation rules ask to discuss about them, if anything, and to what extent GPs are instructed to investigate the sociality of the diagnostic concepts they introduce into the lives of their patients. The next section zooms in on how the role of GPs in Dutch patients’ lives developed to what it is today.

5.2.2 Historical and contemporary roles of Dutch GPs (‘the who’)

5.2.2.1 *Grappling with paternalism in the formative years of medical ‘people specialists’*

This first section in the part of the chapter that describes the role authorities of Dutch GPs (‘the who’) provides relevant historical background to the next part’s discussion of the domain’s explanation paradigm. It describes a group of societally engaged medical practitioners who specialized in a highly personal and paternalistic practice, and the reasons why they decided to change into more reserved and truthful practitioners. The next part connects such developments to legal efforts in support of the move towards ‘informed consent.’

When in 1865 Dutch law established the legally protected title of medical doctor exclusively to doctors with schooling in the natural sciences, most of them were general practitioners by default.⁷¹⁷ By the time access to healthcare for poor and low-income citizens was established in the late 1940’s people flocked to them in large numbers. They presented their doctors with a very broad arrange of ailments whose causes and symptoms transcended the strictly medical, bringing social, societal and psychological subjects into consultations.⁷¹⁸ Although ‘mental health’ had by then established as part of a fast growing set of specialist medical trades, GP’s were reluctant to refer their patients and commenced to expand their own ground.⁷¹⁹ The embrace of the mental and societal dimensions of personal health was part of their gradual self-definition and positioning as societally engaged health care professionals, ‘patient specialists’ in relation to their more physiologically oriented, ‘disease specialist’ peers.⁷²⁰

715 Tamar Sharon, ‘Self-tracking en sociale netwerken in de gezondheidszorg. Verschuivende definities van gezondheid en patiënt-zijn’, in *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, and Gerben Meynen (Boom Filosofie, 2013), 279–80.

716 Schermer, ‘Wat is Ziek, Wat is Gezond?’, 120.

717 Dick Willems, ‘Family Medicine’, in *Encyclopedia of Global Bioethics*, edited by Henk ten Have (Cham: Springer International Publishing, 2014), 1–10.

718 Mol and van Lieshout, *Ziek is het woord niet: medicalisering, normalisering en de veranderende taal van huisartsengeneeskunde en geestelijke gezondheidszorg, 1945-1985.*, 96; Jolanda Dwarswaard, ‘De Dokter en de Tijdgeest’ (Erasmus university, 2011).

719 Mol and van Lieshout, *Ziek is het woord niet: medicalisering, normalisering en de veranderende taal van huisartsengeneeskunde en geestelijke gezondheidszorg, 1945-1985.*, 98.

720 Dwarswaard, ‘De Dokter en de Tijdgeest’, 58.

Throughout the 1950's and 60's GPs developed a highly paternalistic stance. In the image of authoritative fatherhood, conversation (mainlined as a diagnostic tool) was rather one-way: a means to 'get' to knowledge, but not to share it, let alone to share decision making on that basis. The authoritarian stance was eventually embraced as an actual part *of* treatment for the placebo-effect it was regarded to have.⁷²¹ Truthfulness in consultations and diagnostics was no requirement; bending truths was acceptable practice if patients were seen to benefit.⁷²² The worst kind of truths, those of imminent death, were concealed as a rule.⁷²³

For a range of reasons both developments (the strong societal dimension and the authoritative paternalism) eventually became criticized from within. GP treatments lacked (accepted) scientific grounds and common standards, and in combination with the practice's 'cult of personality' this had led to highly divergent practices. Concerns were voiced about credibility *vis-à-vis* other groups of practitioners,⁷²⁴ as well as patients. Students flocked to the 'official' specialist fields whose scientific standing was held in higher regard. And in absence of specialist GP education, those who did choose the GP trade after their basic medical education were badly prepared for what GP practice and its very personal patient relations required.⁷²⁵ Over-confident GPs landed in court for missing important diagnoses of long-term patients whom they assumed too much knowledge about,⁷²⁶ and the fact that patients could become anxious because of a *lack* of information rather than knowledge about their states started to land.⁷²⁷

Efforts to improve these situations led to the establishment of a National GP Council, a specialist journal (*GP and Science*), a specialist internship and eventually, in the early 1970's, a mandatory specialist educational track. The association of GP practice

721 In 1957, French (?) psychiatrist Balint, a proponent of the understanding of this function of a GP had become embraced Esther van Osselen et al, *Geschiedenis van de Huisartsgeneeskunde* (Utrecht: Nederlands Huisartsen Genootschap, 2016), 36,41.

722 Dwarswaard, 'De Dokter en de Tijdgeest', 126.

723 Mol and Van Lieshout cite a GP who warns how patients may 'trick' them into affirming their impending death by falsely mentioning how they knew they would not make it. Mol en van Lieshout, *Ziek is het woord niet: medicalisering, normalisering en de veranderende taal van huisartsgeneeskunde en geestelijke gezondheidszorg, 1945-1985.*, 219.

724 Dwarswaard, 'De Dokter en de Tijdgeest'; See also van Osselen et al, *Geschiedenis van de Huisartsgeneeskunde*, 47, 49.

725 Interview with Bob de Groot, retired psychologist who was present from the start and took part in the design of the curriculum Geurt Essers, 'De huisartsopleiding in Nederland is al 50 jaar een succesformule', *Huisarts en wetenschap* 64, nr. 7 (July 2021): 6-8 The whole interview was published online <https://www.huisartsopleiding.nl/over-de-organisatie/50-jaar-huisartsopleiding-in-nederland/geschiedenis-van-de-huisartsopleiding/>.

726 Willems, 'Family Medicine', 6.

727 Mol and Van Lieshout cite a GP who warns how patients may 'trick' them into affirming their impending death by falsely mentioning how they knew they would not make it. Mol en van Lieshout, 219.

with the University track led to what is described as functional cross fertilization.⁷²⁸ Practicing GPs became patrons of the specializing students, and contributed to their university education where they became part of a team that also included psychologists and other behavioral scientists.⁷²⁹ GP insights came to inform basic medical curricula, and a GP internship was added to the list of obligatory clinical internships that all medical students participated in.⁷³⁰

Several further changes were put in motion from the 70's onward. With regard to treatment, peer-assessment and obligatory National GP Council membership were introduced.⁷³¹ Universities commenced to harmonize the educational tracks of their GP schools. A basic GP task package was agreed on in 1983, and the next millennium saw the development of a national curriculum that described shared goals, competences, and terms in detail. A national education institute was established to govern the separate institutes.⁷³²

With regard to the GP-patient interactions, the 70's and 80's saw a move towards (some) more honesty, and away from the earlier paternalism. The passive, docile, highly dependent patient that had come to unquestioningly rely on their medical authority had become a burden.⁷³³ At a time that patients themselves became more vocal (in line with the spirit of the time) Dutch GPs became proponents of what they saw, or wanted to see, as a more independent type of patient. A new kind of paternalism surfaced: GPs advocated the need to 'educate' their patients into responsible (self-caring), knowledgeable, critical citizens who came to their GPs with a self-defined 'help request'.⁷³⁴

But these ideals of patient autonomy also raised concerns. It was (and is) questioned whether patients could meaningfully reply to the question that was put to them at the start of new-style consultations: "what do you seek my help for?" It was (and is) argued that what patients need help for, is knowledge that is necessarily made through, and not in isolation from, the exposure to a GP's knowledge.⁷³⁵ Additional concerns

728 van Osselen et al, *Geschiedenis van de Huisartsgeneeskunde*, 45–47.

729 Although in the early years students reportedly struggled to deal with their equally authoritative, but differently opined 'masters' (academic educators and their GP patrons whose knowledge was rooted in practice) Author's interview with De Groot and Essers, 'De huisartsopleiding in Nederland is al 50 jaar een succesformule'.

730 van Osselen et al, *Geschiedenis van de Huisartsgeneeskunde*, 51–52.

731 Dwarswaard, 'De Dokter en de Tijdgeest', 112.

732 van Osselen et al, *Geschiedenis van de Huisartsgeneeskunde*, 52–53.

733 Dwarswaard, 'De Dokter en de Tijdgeest'.

734 Mol en van Lieshout, *Ziek is het woord niet: medicalisering, normalisering en de veranderende taal van huisartsgeneeskunde en geestelijke gezondheidszorg, 1945-1985.*, 120.

735 Mol en van Lieshout 40-41.

pertained to how the focus on individual, rather than societal causes could consolidate systemic, societal problems.⁷³⁶

These developments coincided with governmental goals to cut public health expenses.⁷³⁷ GP's efforts towards a more standardized and 'efficient' practice came to serve the dual goals of keeping patients *and* the public health budget healthy. Unnecessary treatment was to be avoided. Earlier traditions to let patients exit a consultation with at least a prescription for e.g., a pain killer were departed from and a large array of more serious treatments were de-prioritized (such as antibiotics for ear and throat infections). GPs explicitly consulted on the premise that the body's own healing power resolved most complaints over a fortnight, and asked their patients to come back if they would not. The 90's saw a further cultivation of the conservative practice that Dutch GPs are seen to have today with regard to prescribing medication, specialist diagnostics, and specialist treatment referrals. There was (and is) however also critique on this strong gatekeeper disposition. Patients are increasingly aware of novel medical developments, and when their knowledge and wishes with regard to treatment and diagnostics are structurally ignored, the doctor-patient relationship suffers.⁷³⁸

5.2.2.2 *General Practitioners today: personal and political relations to maintain and resist*

The last two sections discuss the roles that modern GPs are described to have in citizens' lives, and in society, today. Although GPs increasingly work in group practices, serving modern families whose members don't necessarily share GPs, many GP-patient relations are still long-term and GPs' engagement with the network around their patients is especially large in such cases.⁷³⁹ They know about (and treat) neighbors, children, parents. They are confronted with influences of these relations as well as those of e.g. school, work, personal economy. And they are still confronted with the influence of societal and political developments next to the more strictly medical.⁷⁴⁰

When family relations *are* involved, specific dilemmas with regard to GPs' professional, legal, and ethical duties can arise.⁷⁴¹ GPs may learn something about one family member that is relevant, but yet unknown to another member (genetic dispositions, hereditary

736 Mol en van Lieshout, *Ziek is het woord niet: medicalisering, normalisering en de veranderende taal van huisartsgeneeskunde en geestelijke gezondheidszorg, 1945-1985.*, 224.

737 Especially in comparison to other European countries. van Osselen et al, *Geschiedenis van de Huisartsgeneeskunde*, 122.

738 Dwarswaard, 'De Dokter en de Tijdgeest', 81–85.

739 Willems, 'Family Medicine'.

740 Willems, 5.

741 Willems, 'Family Medicine'.

diseases, heritage itself).⁷⁴² Telling this other person patient might breach that patients' right *not to know* (think of incurable and/or deadly diseases.)⁷⁴³ Information can also be offered to GPs by other persons (a school teacher, a neighbor, an ex-partner), and it can be false. GPs are also relied on to know when to breach their patient's confidentiality and report to public institutions, for example in cases of suspected abuse. They are trained in cross-disciplinary reasoning to deal with all such situations.⁷⁴⁴

GPs also need to maintain relations with all kinds of specialist peers, clinics, insurers, pharmaceutical companies and notably, the State and its administrative bodies. Tensions between them and the State arise as a consequence of conflicting interests and different opinions with regard to the conditions for good patient care. E.g., rising costs because of longer life spans and ever advanced medical possibilities are not germane to the Netherlands, but characteristic tensions are seen to arise there as a result of the inclination of consecutive Dutch governments to adopt ill-advised notions of patient (+citizens!) autonomy, self-sufficiency and 'health literacy'.⁷⁴⁵ But the active, well-informed 'good' patient⁷⁴⁶ is scarcer than calculated for. They are also unequally distributed across the national population. Public GP budget choices based on the Dutch population's 'average needs' is therefore argued to put disproportionate pressure on the quality of GP care in various regions.⁷⁴⁷ Empathy and compassion are seen to be under pressure in such 'market driven' practices, which the field defines as therapeutically problematic.⁷⁴⁸

5.2.2.3 *Medical expert, communicator, collaborator, leader, health advocate, scholar, professional: GPs in the national training curriculum*

This last section discusses some core *competences* that GPs are expected to master in the national GP training curriculum. The competences express important norms with regard to GPs social-epistemic positions and endeavors, also with regard to explanation specifically. As such, they will be referred to in later sections. In the early

742 Wouter De Ruijter, Aart Hendriks, and Marian Verkerk, red., *Huisarts tussen individu en familie: morele dilemma's in de huisartspraktijk* (Van Gorcum, 2012), Cases 17 and 18.

743 H.H.J. Leenen et al, *Handboek gezondheidsrecht*, edited by J Legemaate, 8e editie (Den Haag: Boom Uitgevers, 2020), 126.

744 'GP's in between individual and family' discusses such family conundrums from medical, legal, and ethical perspectives De Ruijter, Hendriks, and Verkerk, *Huisarts tussen individu en familie: morele dilemma's in de huisartspraktijk*.

745 Understood as 'the combination of cognitive and social skills needed to adequately handle information about health, sickness, and care', Twickler et al, 'Laaggeletterdheid en beperkte gezondheidsvaardigheden vragen om een antwoord in de zorg', *Nederlands Tijdschrift voor Geneeskunde* 2009, nr. 153:A250, last consulted 24 June 2020, <https://www.ntvg.nl/artikelen/laaggeletterdheid-en-beperkte-gezondheidsvaardigheden-vragen-om-een-antwoord-de-zorg/volledig>.

746 Legemaate, 'Nieuwe Verhoudingen in de Spreekkamer: Juridische aspecten', 16.

747 Jany Rademakers and Nederlands instituut voor onderzoek van de gezondheidszorg (Utrecht), *Kennissynthese: gezondheidsvaardigheden: niet voor iedereen vanzelfsprekend* (Utrecht: NIVEL, 2014).

748 van Osselen et al, *Geschiedenis van de Huisartsgeneeskunde*, 124.

2000's: the (Canadian) CanMeds Model⁷⁴⁹ was introduced in all medical Bachelor's and Master's programs, and in Dutch GP specialist education. The model describes six areas of competence and related capabilities that the 'competent medical expert' is expected to master. Per the Dutch curriculum, "a doctor should not only be a medical expert, s/he has to be a communicator, a collaborator, a leader, a health advocate who acts in society's interests, a scholar who thinks in scientific and moral-ethical terms, and a professional who shares knowledge, attitude and skills with others."⁷⁵⁰ GP-specific descriptions for the competence areas together form the competence *profile* that GPs need to make their own. The competence 'communication' places much emphasis on establishing a responsible, trusting working relationship conducive to 'constructive dialogue.' A strong personal engagement with patients, their backgrounds, and environments is needed to bring out their needs and values, and to verify their responsible understanding.⁷⁵¹ The 'communication' competence is part of most of the 10 *themes* that GP education is now organized around. The themes are a recent educational reform. It implements the insight that training for the themes needed to become explicit rather than implied to serve 'future proof' GP practice. The themes are designed around different denominators such as patient types (e.g., youth, patients with unexplained complaints), care types (e.g., urgent care, complex comorbidity), aims (e.g., prevention) and organization (e.g., tuning into a practice's patient population's information needs.) The Characteristic Professional Activities (CPA's) that are described for each theme are especially explicit in how they describe typical combinations of competence-related capabilities.⁷⁵² The CPA's allow to train for observable (and therewith assessible) actions and behaviors, guiding students in their practical and context-specific development.⁷⁵³ The new system is expected to do more justice to GP practice in which different competences are generally combined in the performance of individual activities, and become meaningful in different ways in different cases. The themes and CPA's will be revised every three years to align with scientific, societal, and practice developments. In later sections, some CPA's will be cited around explanation-related subjects.

5.2.3 EBM and SDM: making two core paradigms work for GP practice ('the how')

5.2.3.1 Introduction

This last entry in the 'what, who, & how' part of the chapter discusses two important concepts that have been paradigmatic for medical decision making overall: Evidence Based Medicine (EBM) and Shared Decision Making (SDM). Both 'systems' also

⁷⁴⁹ <http://canmeds.royalcollege.ca/>.

⁷⁵⁰ <http://canmeds.royalcollege.ca/>, see also 'Raamplan Medical Training Framework'.

⁷⁵¹ Huisartsopleiding Nederland, *Competentieprofiel van de huisarts*, 2016, https://www.huisartsopleiding.nl/images/opleiding/Competentieprofiel_van_de_huisarts_2016.pdf.

⁷⁵² <https://www.huisartsopleiding.nl/opleiding/thema-s-en-kba-s>.

⁷⁵³ Huisartsopleiding Nederland, *Competentieprofiel van de huisarts*.

raise specific challenges for, and produce functional discussions about GP practice in particular. The domain's explanation paradigm cannot be meaningfully discussed without some basic knowledge of them both. The two topics were selected on these merits, but also because the important subjects and themes come in via the discussions they raise; subjects that would otherwise need to be dealt with separately in what is already a 'heavy' chapter section.

This is specifically true for the subject of medical technology: instruments and materials, medication, software. The complexity and multi-disciplinarity of medical technologies places GPs in a vast field of knowledge making networks. This brings in interesting questions around things like interdisciplinary trust⁷⁵⁴ and the furthering of shared medical knowledge spheres.⁷⁵⁵ E.g., GPs are required to keep up with revelations and publications on *untrustworthy* artifacts, medicine, and (digital) methods.⁷⁵⁶ This has not become easier with the advance of AI.⁷⁵⁷ Consumer health technologies already pose growing challenges for GPs in this regard. These produce various kinds of 'measurements,' and even diagnostic suggestions.⁷⁵⁸ The perceived 'lagging' uptake of such developments by Dutch GPs is qualified in different ways, depending on who's talking: as cautious, conservative, harmful to the developing market, and/or an obstacle to the improvement of GP care and patient autonomy.⁷⁵⁹ All these subjects and themes will come up again over the following sections.

754 Sophie van Baalen and Annamaria Carusi, 'Implicit Trust in Clinical Decision-Making by Multidisciplinary Teams', *Synthese* 196, nr. 11 (1 November 2019): 4469–92.

755 Which also makes them dependent on the other disciplines' self-understanding and the ability to explain; not to mention challenges of interdisciplinary assessment - a problem not at all germane to medical knowledge making of course, as this interesting discussion of interdisciplinarity in scientific practice in general shows: Katri Huutoniemi, 'Interdisciplinarity as Academic Accountability: Prospects for Quality Control Across Disciplinary Boundaries', *Social Epistemology* 30, nr. 2 (2016): 163–85.

756 An effort that goes beyond national borders in significant ways; e.g., 'things' are produced and distributed internationally but held to different standards in different places. With regard to medical artifacts, EU governance has for example been found lacking compared to the US, illustrated among other things by the [large numbers] of patients in the EU who suffered the consequences of harmful breast and hip implants. To improve this situation, and at the same time expand the governance of increasingly 'AI-driven' medical decision support systems, the EU Medical Devices Directive was recently replaced by the Medical Device Regulation (MDR).

757 The progress of (inscrutable) AI in this dimension is tabled as a complicating factor of doctors' legal, ethical, and moral end-responsibility 'Digitale dokters: Een ethische verkenning van medische expertsystemen - Signalement - CEG – Centrum voor Ethiek en Gezondheid' (Centrum voor Ethiek en Gezondheid, 2018); Pierce, Sterckx, en Van Biesen, 'A Riddle, Wrapped in a Mystery, inside an Enigma'; Powles en Hodson, 'Google DeepMind and Healthcare in an Age of Algorithms'.

758 These are developed and produced outside of the medically regulated research and production spheres posing obstacles to GP's assessment of them, but they allow such tech to play a role in treatment (as they are increasingly asked to do by their patients) they will be responsible for their effects. Consumer health's problematic nature is researched in other spheres, too, saliently in consumer protection law, for example by Sax: Sax, 'Optimization of What?'.

759 Anna V. Silven et al, 'Clarifying Responsibility: Professional Digital Health in the Doctor-Patient Relationship, Recommendations for Physicians Based on a Multi-Stakeholder Dialogue in the Netherlands', *BMC Health Services Research* 22, nr. 1 (2022): 129.

5.2.3.2 Evidence Based Medicine: merits, critiques, and alternatives

The start of EBM is generally pinned to the early 1990s. Dutch medical ethicist, and long-time GP educator Willems describes EBM's birth as a 'point on a path' marked by various other and earlier developments; a path that pushed medicine forward from practice to science.⁷⁶⁰ Anatomy and pathology at the turn of the nineteenth century, experimental and statistical methods in the mid 1800s, advances in pharmacology in the 1940s and the standardization of scientific reporting that followed from the internationalization of medicine in the years after that.⁷⁶¹ Two key publications are linked to the start of EBM: Archie Cochrane's 1972 *Effectiveness and Efficiency: Random Reflections on Health Services and the Dartmouth Atlas of Health Care*.⁷⁶² Cochrane argued to engage with three lines of questioning in the assessment of medical decisions: theoretical (plausibility), effectiveness (empiricism), and proportionality (risks, costs, burdens). The Dartmouth Atlas charted the huge variation in medical practices across the US, revealing it to be problematic. A much-named example from it is the persistence of radical mastectomies after ample evidence that less intrusive surgery could be just as effective.⁷⁶³ In The Netherlands as well the call for better justified medical decision making had become louder. Patients were treated unnecessarily; clinical decisions suffered from discriminatory bias, subjectivity, and uncertainty; responsible patient trust was seen to suffer.⁷⁶⁴ EBM was looked to for improvement through the use of statistical evidence, rule-based reasoning, the creation of standards based on (summaries of) scientific publications, and the reduction of the complexity of medical practice through protocollization and standard setting. Administrative bodies and insurers, for their part, hoped that the introduction of EBM would help to reduce Public Health Care expenses that had been growing rapidly for decades.

The concept developed into a dominant paradigm across medical domains. And although EBM's beneficial influence is acknowledged, its premises are also pulled into doubt and some of its effects are deplored. For one, the quantitative methods were no

760 Dick Willems, 'Bewijzen, weten, en begrijpen. Drie vormen van kennis in de zorg', in *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, and Gerben Meynen (Boom Filosofie, 2013); D.W. Willems et al, 'Passend bewijs. Ethische vragen bij het gebruik van evidence in het zorgbeleid', Signalement (Centrum voor Ethiek en Gezondheid, 2007), 15–17.

761 Willems, 'Bewijzen, weten, en begrijpen. Drie vormen van kennis in de zorg'; Willems et al, 'Passend bewijs. Ethische vragen bij het gebruik van evidence in het zorgbeleid', 15–17.

762 For the Dartmouth Atlas see <https://www.dartmouthatlas.org/>. Much cited is also Sackett's 1996 definition "The conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."

763 Earlier, consensus-base standardization efforts in physicians' 'closed' circles had (unsurprisingly) failed to establish sufficient progress. Dutch GP and health professor Burgers recounts how the establishment of their 'recommendations' became known as the GOBSAT method, i.e. Good Old Boys Sat Around the Table. J. Burgers, 'Oratie: Persoonsgerichte zorg en richtlijnen: contradictie of paradox?' (Oratie, Maastricht, 2017), <https://cris.maastrichtuniversity.nl/en/publications/persoonsgerichte-zorg-en-richtlijnen-contradictie-of-paradox>.

764 Willems, 'Bewijzen, weten, en begrijpen. Drie vormen van kennis in de zorg'.

cure for discriminatory bias nor ‘Big Pharma’s’ commercial interests.⁷⁶⁵ On a more theoretical level, critique pertained to the promotion of evidence as the only valid form of justification,⁷⁶⁶ to the prioritization of scientific methods over other validators of evidence,⁷⁶⁷ and of quantitative science as representative of all of science (inviting the natural sciences’ notions of ‘proof’ into a field that it cannot be of use for).⁷⁶⁸ Important dimensions of medical practice for whom EBM’s rule-based, standard-driven practice were a bad fit were seen to be at risk.⁷⁶⁹ This was of special concern for GP practice. Neither causes nor treatments for typical GP patients’ ailments were scientifically well understood: ‘common’ colds, aches, bumps, rashes, and other phenomena.⁷⁷⁰ GP practice was also seen to thrive on a strong qualitative approach. It was argued that in order to *responsibly* assess and apply scientific insights (and EBM’s derived standards and protocols), GPs need to be able to make their intimate, often long-term patient experience and patients’ own voices count.⁷⁷¹

Various initiatives meant to push back on these pitfalls. A 2007 report of the Dutch Center for Ethics and Health warned how the implementation of EBM’s systematics posed a risk for the dimensions of attention, trust, and presence, impoverishing a comprehensive understanding of what ‘caring, medically’ entails.⁷⁷² In 2017, the Dutch Council for Public Health and Society voiced concerns about uncritical use of practice guidelines.⁷⁷³ For the GP context in particular, Dutch GP’s argued to ditch the Acronym EBM for EIP, *Evidence Informed Practice*,⁷⁷⁴ or promoted ‘real EBM.’

765 Citing Greenhalgh et al (2014 ADD CITE), Baalen and Boon, ‘Evidence-Based Medicine versus Expertise’, 2.

766 Willems, ‘Bewijzen, weten, en begrijpen. Drie vormen van kennis in de zorg’.

767 Sophie van Baalen and Mieke Boon, ‘An Epistemological Shift: From Evidence-Based Medicine to Epistemological Responsibility’, *Journal of Evaluation in Clinical Practice* 21, nr. 3 (2015): 433–39.

768 Burgers, ‘Oratie: Persoonsgerichte zorg en richtlijnen: contradictie of paradox?’, 7.

769 van Baalen and Boon, ‘An Epistemological Shift’; Van Baalen promotes the integration of different knowledges in clinical reasoning: to understand what that takes, how to assess it, and improve it. She argues that salient EBM critiques stop short of taking a that necessary step, and in addressing this gap, her trail eventually leads to ‘epistemic responsibility,’ and to Lorraine Code’s work. Code’s (and others’) notions are used to describe how responsible doctoring requires the mastering of ‘integrative skills,’ the quantitative and the qualitative, and becoming conscious of their interplay. Sophie Jacobine van Baalen, ‘Knowing in Medical Practice: Expertise, Imaging Technologies and Interdisciplinarity’ (2019), <https://research.utwente.nl/en/publications/knowning-in-medical-practice-expertise-imaging-technologies-and-in>.

770 Burgers, ‘Oratie: Persoonsgerichte zorg en richtlijnen: contradictie of paradox?’

771 Jean Muris, Roger Damoiseaux, en Nynke van Dijk, ‘Bijwerkingen en valkuilen van EBM: trap er niet in!’, *Huisarts en wetenschap* 60, nr. 11 (November 2017): 548–51.

772 Willems et al, ‘Passend bewijs. Ethische vragen bij het gebruik van evidence in het zorgbeleid’.

773 The report was also published in English ‘No evidence without context. About the illusion of evidence-based practice in healthcare’, publicatie (Raad voor Volksgezondheid en Samenleving, 2017).

774 Glaziou cited by Willems Willems, ‘Bewijzen, weten, en begrijpen. Drie vormen van kennis in de zorg’, 198.

a critical, standardization-resistant GP practice.⁷⁷⁵ Concerns were voiced about which (if any) norms, values, principles and choices from the ethical, qualitative GP realm were incorporated in EBM-based practice guidelines⁷⁷⁶ and decision making aids,⁷⁷⁷ and that GPs need to embed their use in qualitative practices.⁷⁷⁸ The Dutch Health Care Authority advised to collaborate with patient organizations and focus groups around their creation.⁷⁷⁹

One type of ‘qualitative knowledge making’ promoted for (and practiced by) GPs are moral consultation methods. Different from medical ethics that aim to sustain decision making more directly, moral consultations mean to train reflective capabilities as a goal in itself.⁷⁸⁰ Which still supports decision making, of course.⁷⁸¹ But the focus is questions rather than answers, and consensus is not the aim. The methods broadly fall into two types; ‘problem-oriented’ and ‘position-oriented,’ where the latter specifically helps to address larger questions on what good care is relative to specific challenges or actors.⁷⁸² A variation that includes patients and their (other) (in)formal carers is described as a “structured exercise in shared learning and exploring.”⁷⁸³ Especially promoted for ‘hard cases’ (think of treatment refusals or the opposite, not wanting to know about hereditary

775 Jean Muris, Roger Damoiseaux, and Nynke van Dijk, “Bijwerkingen en valkuilen van EBM: trap er niet in!,” *Huisarts en wetenschap* 60, no. 11 (November 2017): 548–51.

776 D. Willems en M. Hilhorst, *Ethische problemen in de huisartspraktijk*, Practicum Huisartsgeneeskunde (Bohn Stafleu van Loghum, 2016), 101; Not so Burgers, who argues that ethical values such as *in dubio abstine* and *primum non nocere* are generally referred to in guidelines, which makes then responsible instruments to use that would not be used by GP’s if they weren’t. Burgers, ‘Oratie: Persoonsgerichte zorg en richtlijnen: contradictie of paradox?’, 10.

777 E.g. a breast cancer treatment decision aid based on ‘patient reported experience/outcome measures’ (PREMS/PROMS) did not include concerns with regard to ‘ability to perform professionally,’ but did include concerns of ‘body contour.’ An omission with dire consequences when one considers that some contour-saving treatments lead to much longer, or even (partial) permanent labour incapacity depending on the patient’s profession. Ingeborg Engelberts, Maartje Schermer, en Awee Prins, ‘Een goed gesprek is de beste persoonsgerichte zorg’, *Medisch Contact* 2018, nr. 28/29, last consulted 16 July 2020, <https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/een-goed-gesprek-is-de-beste-persoonsgerichte-zorg.htm>.

778 Willems, ‘Bewijzen, weten, en begrijpen. Drie vormen van kennis in de zorg’, 200.

779 Burgers, ‘Oratie: Persoonsgerichte zorg en richtlijnen: contradictie of paradox?’, 15; See also: Willems en Hilhorst, *Ethische problemen in de huisartspraktijk*, 101; Health law scholar Leegemate however warned that the move to include patients, and not just GP’s in guideline creation was preceded by decades of concerns of patients, politicians, and others Legemate, ‘Nieuwe Verhoudingen in de Spreekkamer: Juridische aspecten’, 10.

780 “Moral consultations can neither replace medical ethics’ roles in solving medical ethical dilemmas, nor replace it in creating ethics policies/guidelines.” Hans van Dartel and Bert Molenwijk, red., *In gesprek blijven over goede zorg: Overlegmethoden voor moreel beraad* (Boom Filosofie, 2014), 11–15, 90–95.

781 Dick L. Willems, ‘Ethiek en de huisarts’, *Bijblijven* 32, nr. 3 (1 April 2016): 138.

782 Willems, ‘Ethiek en de huisarts’.

783 van Dartel and Molenwijk, *In gesprek blijven over goede zorg: Overlegmethoden voor moreel beraad*.

conditions and pregnancy), this (variant of the) method allows parties to discuss their individual as well as applicable legal, ethical, and professional norms.⁷⁸⁴

Support for qualitative approaches also expresses in GP education materials. Reflective practices and attitudes are trained for, patients' individual and collective epistemic positions need to be engaged with. An instruction about cases of badly understood ailments⁷⁸⁵ (relevant also in light of the remark about the commonness of such ailments in GP practice) is illustrative. The training booklet instructs that in absence of known *causes* to treat, the trainee-GP is prompted to remain upfront about their medical uncertainty ("not act surer than they are"), to take their patient's own interpretations seriously (and be prepared to refer for the specialist care they seek), keep looking for somatic clues⁷⁸⁶ and work *with* them towards a *shared* understanding of how to best deal with the symptoms.

The last two paragraphs brought in the salient other of the GP: the patient, the person whose meaningful participation in decision making is required. The next section takes a closer look at the concept of SDM and the challenges it comes with—including those posed by EBM.

5.2.3.3 *Harmonizing doctor and patient knowledge in Shared Decision Making: aims and challenges*

The point of SDM, simply stated, is that patients are enabled to bring their "expectations, goals, preferences, and values" to bear in medical decision making: to make sure that these are part of the translation of scientific medical knowledge to their individual states and situations.⁷⁸⁷ For this, doctors and patients need to *exchange* information. Simply stated again, patients need medical information before they can voice a preference, and doctors will need to know about patients' goals to understand and explain the value of any particular treatment for them. In practice, this is no simple task. The collaborative knowledge making that SDM practice requires needs a social-

784 Johan Legemaate en Guy Widdershoven, red., *Basisboek ethiek en recht in de gezondheidszorg* (Boom Filosofie, 2016), 216; Patient participation is also discussed by Weidema et al in van Dartel and Molenwijk, *In gesprek blijven over goede zorg: Overlegmethoden voor moreel beraad*.

785 A Characteristic Professional Activity ('KBA') in the theme of Unexplained Physical Complaints ('SOLK' in Dutch), Theme 7 in GP training materials Huisartsopleiding Nederland, *Thema's en KBAs*, 2016.

786 An important instruction since medical literature is rich with examples where the 'vague complaints' of patients from groups that are not taken seriously are wrongly dismissed as psychological or psychosomatic, resulting in missed diagnoses and fatalities. On 28 February 2022, The Municipality of Amsterdam announced this as a focus point of an array of measures to improve the quality of care for marginalized groups. David Hielkema, 'Gediscrimineerde medewerkers en eenzijdige blik bij diagnoses: Amsterdam gaat racisme in de zorg aanpakken', *Het Parool*, 28 February 2022, sec. Amsterdam, <https://www.parool.nl/gs-bb761baf>.

787 Pierce, Sterckx, and Van Biesen, 'A Riddle, Wrapped in a Mystery, inside an Enigma'.

epistemic relationship conducive to that aim, and medical history teaches how such relationships are neither easily, nor naturally established.

The shift from authoritarian decision making to the inclusion of patients in deliberations shares important themes with the shift from ‘sanctioned deceit’⁷⁸⁸ to the informed consent paradigm. Both revolutions involved challenges to physicians’ near-absolute dominance in both knowledge and decision making. But although the two shifts also share much historical time, the informed consent explanation obligations were taken up more easily (but not easily) than notions of shared decision making. The quality of explanations that were given before SDM was embraced suffered as a consequence. For the Dutch GP context, it is argued that SDM finally brought about what informed consent ‘proper’ meant to establish all along.⁷⁸⁹ The following paragraphs first provide historical background that makes insightful how the end of the ‘doctor knows best’ era did not mean the end of authoritative decision making, then zooms in on contemporary GP practice.

Section 5.2.2 discussed how honesty and non-persuasiveness were promoted in a move away from highly paternalistic Dutch GP practices at the end of the 1960’s. But GPs also reported how their patients were uncomfortable with honest conversations, and seemed distrustful when they were sent on their way without at least a prescription.⁷⁹⁰ In *The Silent World of Doctor and Patient*, Katz places complaints about patients’ unwillingness to exit the authoritarian doctor-patient relationship (which were also much heard in the US) in critical historical perspective. In his account, patients’ purported lack of emotional and cognitive capacity and capabilities to deal with bad news and medical information were an excuse. It was doctors themselves who were reluctant to ‘let go’ after centuries of cultivating hierarchical, authoritarian practices,⁷⁹¹ and they themselves were to blame for the lack of medical ‘capabilities’ of their patients. For centuries, patients had been conditioned to accept authority as beneficial if they even had such a choice (most of them did not) and to understand *themselves*

788 Katz, *The Silent World of Doctor and Patient*. 1984, Katz’s discussion of the social power dynamics of medical knowledge making communities recounts how a long history of (all kinds of) doctors’ near absolute social-epistemic authority vis-a-vis patients correlated with doctors’ own lack of ‘scientifically sound’ understanding of the why’s and how’s of sickness, disease, and cure. The result, in his words, was ‘sanctioned deceit.’

789 J. Legemaate, ‘Rechtstekorten in het gezondheidsrecht’, *Tijdschrift voor Gezondheidsrecht* 42, nr. 3 (2018): 202.

790 Dwarswaard, ‘De Dokter en de Tijdgeest’, 126.

791 Katz discusses how reasons for excluding patient influence vary across domains, time, and patient populations, and contrary voices do pop up every now and then—but they don’t win. E.g., the unfreedom of poor, enslaved, and uneducated people precluded their right to medical information for various reasons in different times, such as that medical information was unuseful for people who had no medical choices to make if they could understand. Listening to patients reserved for those who were seen to be able to articulate what ailed them in any usable quality. And when they were deemed ‘worthy’ conversation partners, they were still not invited to decision making: their freedom of expression simply lead to better treatment.

as ignorant, medically. No-one had offered them the thinking tools they needed now.⁷⁹² This also features in much-heard complaints that patients did not understand ‘technical details,’ and therefore had no use for medical knowledge. In fact, argues Katz, it had never been investigated what aspects of medical knowledge were useful at all to know for patients, and would help them make decisions. This was not just true for technically complex knowledge, but for all medical situations: the ‘cult of silence’ had been complete.⁷⁹³ By characterizing medical knowledge as technical, attention was also diverted from doctors’ own proper understanding.⁷⁹⁴ An understanding that suffered from a hierarchical medical culture not conducive to honesty and reflectivity. Medical controversies (and there were many) weren’t used to learn from, and persisted for long periods of time in absence of cross-rank critique.

In light of this history, and in such a culture, Katz argued that relying on ‘doctor knows best’ is as dangerous as trusting that doctor and patients share ‘identity of interests.’⁷⁹⁵ Both things cannot be established without proper exchanges. And for patients to give meaningful input, they need honest and meaningful information. When this does not happen, the unhealthy hierarchical doctor-patient relation (indeed) persists. He argued that a range of virtuous activities should be(come) obligatory to make the informed consent obligations into meaningful guides to proper, meaning shared, decision making.⁷⁹⁶ Doctors need to acknowledge the limitations of their knowledge and of medical knowledge in general, to understand that communicating medical knowledge needs is a tailor-made practice, and to (learn to) trust their patients’ reflective and emotional capabilities.⁷⁹⁷

792 Katz’ also notes how much patients’ behaviour is wrongly qualified as ‘trusting’ to begin with: not engaging in deliberation, discussion, and ‘honest exchanges’ may also signify distrust. He names the example of non-compliance with treatment prescriptions. This part of Katz account leans heavily on Freudian and other psycho-analytical theory that was highly popular at that time, and is left aside here.

793 Katz, *The Silent World of Doctor and Patient*. 1984, 93.

794 His citation from the French Doctor/Priest/Philosopher Samuel de Sorbière’s (mildly ironic) ‘*Advice to a Young Physician Respecting the Way in Which He Is to Conduct Himself in the Practice of Medicine, in View of the Indifference of the Public to the Subject, and Considering the Complaints that Are Made about Physicians*’. Provides some comic, albeit cynical relief. De Sorbière imagines a young doctor being honest to his patient: “Although I am disposed to be of service to you, and will undertake your cure as an end to be hoped for, and with God’s help achieve some success, in order to safeguard your interests, I must tell you that medicine is a very imperfect science, that is quite full of guesswork, that it scarcely understands its subject matter, nor is it familiar with the things employed to maintain it; that the more enlightened only feel their way in it groping amidst a thick gloom; and that after having considered seriously all the matters which may be useful, collected all one’s thoughts, examined all one’s experiences, it will indeed be a wise physician who can promise relief to a poor patient.” Katz, 10–11.

795 “in the absence of any one clear road to well-being, identity of interests [between doctor and patient] cannot be assumed, and consensus on goals, let alone on which paths to follow, can only be accomplished through conversation.” Katz, xlv.

796 Katz, 102.

797 Katz, 102.

In the studied literature on the Dutch GP domain, a much-named goal, or necessity for sharing decisions with patients responsibly is the growing number of independently living elder citizens. To support them in ‘managing’ their various chronic states, ever growing ranges of treatment and self-care options need to be discussed, monitoring agreements made.⁷⁹⁸ Also mentioned however are how from the late 1980’s onward, Dutch Public Policy started to treat citizens as more self-sufficient, autonomous ‘consumers’ of public services than they in fact were, which was especially problematic for citizens who are dependent on public services—a given in the medical domain.⁷⁹⁹ Enabling patients to participate responsibly means understanding that they do not possess a clear and ‘autonomous,’ stable staple of concerns, wishes, preferences, and values. Where this is a scientifically shaky premise in general,⁸⁰⁰ all these things are also known to be in flux after the confrontation with a fundamentally impactful disease.

This has consequences for the capabilities that GPs need to engage in responsible SDM: they need the qualitative skills and methods to help them identify their patient’s values through conversation.⁸⁰¹ All kinds of ‘non-medical,’ social-relational considerations may influence their patients’ preferences. E.g., for elderly persons, feelings of redundancy, of costliness to society, of a fulfilled life, of family and (informal) care-related considerations may be at play, as well as (other) cultural considerations.⁸⁰² The need to be alert to such dimensions, and how that also takes “less objective” assessment methods is explicit in the GP education theme on ‘elderly patients with complex problems.’⁸⁰³

But as the previous discussions of patients’ own lack of medical knowledge, and their sensitization to authoritarian relationships with their doctors forewarned, they too need ‘capability training.’ Health law scholar Legemaate argues that if SDM is to deliver on

798 See e.g. ‘Handreiking Samenwerking huisarts en specialist ouderengeneeskunde: Samenhangende geneeskundige zorg voor patiënten met een complexe zorgbehoefte’ (LHV en VerenSo, 2020); ‘Toekomstvisie 2012-2022 | NHG’, last consulted 27 January 2021, <https://www.nhg.org/toekomstvisie>.

799 Whether treating patients as consumers serves them in terms of getting the guidance they is also critiqued per se. For a fundamental critique on ‘the logic of choice’ as opposed to ‘a logic of care’, see for example Annemarie Mol, *The Logic of Care: Health and the Problem of Patient Choice* (Routledge, 2008).

800 Especially in the ‘hands’ of autonomy-as-individual freedom based political climates Beate Roessler, *Autonomy: An Essay on the Life well-lived* (Polity, July 2021).

801 Engelberts, Schermer, en Prins, ‘Een goed gesprek is de beste persoonsgerichte zorg’; Skills that they argue don’t simply follow from patient communication training. See also Carson, who wrote that wrote, patients “may not be able to articulate [their truths, values] very clearly, or even identify them with much precision, but they bring their lives and their needs (...) one of the central moral challenges doctors must take up is that of helping sick people to “find their voices.” Ronald A. Carson, ‘Medical Ethics as Reflective Practice’, in *Philosophy of Medicine and Bioethics: A Twenty-Year Retrospective and Critical Appraisal*, edited by Ronald A. Carson and Chester R. Burns, Philosophy and Medicine (Dordrecht: Springer Netherlands, 1997), 12.

802 Willems en Hilhorst, *Ethische problemen in de huisartspraktijk*, 10.

803 Huisartsopleiding Nederland, *Thema’s en KBA’s*, theme 4.

the promise to elicit patient values that remained under-recognized before, its methods will also need to elicit knowledge about the capabilities patients need to ‘stand their ground,’ i.e. to participate freely and responsibly, and make themselves understandable in consultations. He suggests policy could be created to support patients in their roles, such as providing them with coaches and guides.⁸⁰⁴ Physician and Health Law scholar Watson additionally warns how social power asymmetries make patients bad self-representatives, and that legal SDM obligations should take care not to ignore, or obfuscate what it realistically takes to establish responsible (and necessary) therapeutic trust.⁸⁰⁵

These considerations are of interest with regard to what the domain’s legal explanation paradigm understands as ‘meaningful positioning’ of the explanation partners. The relation is governed in contract law; a principally horizontal instrument that establishes duties of both parties. Its article 7:452 obliges the patient to provide their doctor, i.e. their contractual partner, with the information and co-operation that allows their doctor to perform their contractual obligations.⁸⁰⁶ The rule was critiqued from the Law’s enactment in 1994 on medical-moral grounds, as well on the basis of its non-enforceability.⁸⁰⁷ But also for missing a fundamental medical point: how can non-medically schooled patients possibly know what information their doctor needs?⁸⁰⁸ With that, the chapter arrives at part two: the discussion of the legal explanation paradigm.

5.3 ‘Information duties’ in the Medical Treatment Agreement Act (‘WGBO’)

5.3.1 Introduction

The second half of the chapter discusses the law’s address of explanation in the GP domain. It reports on an investigation of the field’s main legal explanation obligations, those of the Medical Treatment Agreement Act, WGBO hereafter. Its rules need to be understood in their relation to the larger governance context. Like all behavioral norms in this domain, norms with regard to explanation are established through public and self-regulation: law, medical and professional ethics, professional standard setting. The hierarchical relations between these fields are dynamic and intense discussions

804 Legemaate, ‘Nieuwe Verhoudingen in de Spreekkamer: Juridische aspecten’.

805 Kenneth Watson, ‘Goede zorg, informed consent & shared-decision making: nieuwe basis onder goed hulpverlenerschap en medische aansprakelijkheid?’, *Letsel & Schade* 2018, nr. 3 : 33.

806 This needs explanation: as will be explained in the legal section, the legal relationship of doctors and patients in the Netherlands is shaped in private law’s contractual terms.

807 Leenen et al, *Handboek gezondheidsrecht*, 112, 113.

808 Leenen et al, 113; Johan Legemaate, ‘Patiëntveiligheid en patiëntenrechten (2006)’, in *Oratiebundel Gezondheidsrecht: verzamelde redes 1971-2011* (Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012), 394.

between different norm setters are part and parcel of this structure.⁸⁰⁹ This first section explains the structure of reporting.

The report is structured in two parts. The first part (sections 5.3.1-5.3.2) introduces the WGBO and its recently updated explanation rules. It places the law in relation to the other governance modalities, and addresses several power-sharing concerns of influence on the establishment, interpretation and progress of legal (explanation) obligations.

The second part (sections 5.3.3-5.3.4) discusses main *purposes* of the WGBO's explanation regime. The choice to let these purposes, and not e.g., the elements of the WGBO's legal provisions lead the discussion does justice to the medical field's 'always on' relations of self-standing (explanation) values and their therapeutic ends.⁸¹⁰ E.g., informing patients about known treatment risks serves their informed consent, but the honest conversation this requires also serves the doctor-patient trust relationship and therewith the quality of therapy. There is however not 'one view' from practice, and there are authoritative struggles between the regulating fields. The Lawmaker's choices in the WGBO are therefore embedded in illustrations from the other fields to show how these inform, support, or are in tension with the WGBO's rationales. This is the second reason to choose this reporting structure. By placing Law's choices in this critical commentary, the structure supports the modeled duties of explanation care analysis in part three of the chapter.

The purposes are divided over two main categories. Each is dedicated to one of two sides, or dimensions, of the main historical legal purpose of medical informed consent. This purpose can be summarily described as to secure individual patients' self-determination by protecting their bodily integrity and enabling their decisional autonomy.⁸¹¹ In and outside of law, autonomy and self-determination are typically related to individual freedom in two, related ways: as freedom *from* coercive practices ('negative' freedom), which expresses in the WGBO's aim to 'protect patients from the risk that they cannot self-determine.' The freedom *to* make one's own choices ('positive' freedom) is more attentive to the need for enabling capabilities and

809 Anne Ruth Mackor, "Rechtsregels En Medische Richtlijnen. Een Rechtsfilosofisch Perspectief Op de Aard En Functie van Regels," in *Medische Aansprakelijkheid*, ed. S. Heirman, E.C. Huijsmans, and R. Van den Munckhof, Kenniscentrum Milieu En Gezondheid (Wolf Legal Publishers, 2016); Lena Wahlberg and Johannes Persson, "Importing Notions in Health Law: Science and Proven Experience," *European Journal of Health Law* 24, no. 5 (November 10, 2017): 565–90; N B A T Janssen et al., "Under What Conditions Do Patients Want to Be Informed about Their Risk of a Complication? A Vignette Study," *Journal of Medical Ethics* 35, no. 5 (May 1, 2009): 276–82.

810 Pierce, 'Medical Privacy: Where Deontology and Consequentialism Meet'; see also Johan Legemaate, 'The Development and Implementation of Patients' Rights: Dutch Experience of the Right to Information Patients' Rights', *Medicine and Law* 21, nr. 4 (2002): 731–32.

811 Legal principles and fundamental rights in Health Law, The right to information Leenen et al, *Handboek gezondheidsrecht*, Sections 1.3.2, 2.5.

circumstances.⁸¹² It is this dimension that the WGBO's SDM informed updates meant to serve, and so, these will be discussed in the second category. In both sections, critical views with regard to the mainlining of a strong individualist perception of patient autonomy are included.

5.3.2 The WGBO's explanation regime in the larger governance landscape

5.3.2.1 Power sharing conundrums

After a long period of leaving much medical norm setting up to medical, ethical, professional fields, the development of Dutch Health Law picked up pace from the second half of the 20th century onwards.⁸¹³ And although the general terms of law's codifications are still generally explained, applied, and further developed on the basis of the field's self-set standards, law's ambition is to further medical practices in different ways: by securing established rights and obligations (better) whose uptake is found to be lacking, or, taking a less 'defensive' stance, by promoting societal developments that aren't (sufficiently) taken up in and by the other norm setting fields.⁸¹⁴

Questions about the extent that law *should* codify the field's own norms, and if so to what detail, are part of what arguably will always be a permanent conundrum. Medical/ethical notions and practices thrive on progressive understanding, and law's codification should not arrest this development. But too much legal abstinence makes public governance moot: it lets the field do as it pleases within the open norms of general legal rules.⁸¹⁵ For explanation, law's generality and abstinence turned out to be a problem. Along with initiatives to further implementation, the rules have been updated with more detailed obligations.

A recent Dutch example illustrates how power sharing challenges can surface acutely, leaving medical professionals shy to act. In the heart of the COVID-19 pandemic storm, the scarcity of IC beds necessitated to make 'black scenario' triage rules. The (expected) scarcity was such that no legal, medical or ethical notions could meaningfully determine which patients' care should prevail over others. A public

812 For specific explorations of protecting and enabling free choice in contemporary, commercial digital health environments, see for example Marjolein Lanzing, "'Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies', *Philosophy & Technology*, 6 June 2018, 1–20; Sax, 'Optimization of What?'

813 Leenen et al, *Handboek gezondheidsrecht*, section 1.1.1.

814 This aspect will be dealt with in detail later on. In general, zie e.g. Leenen et al, section 1.1.1; J. Legemaate, 'Aanpassingen van de WGBO', *Tijdschrift voor Gezondheidsrecht* 42, nr. 6 (December 2018): 556–64.

815 Mackor, "Rechtsregels En Medische Richtlijnen. Een Rechtsfilosofisch Perspectief Op de Aard En Functie van Regels"; As mentioned, these discussions are not new, see e.g. on this particular point, J Legemaate, *Goed recht: de betekenis en de gevolgen van het recht voor de praktijk van de hulpverlening*, Preadvies Vereniging voor Gezondheidsrecht (Utrecht, 1994), 46.

debate unfolded. Medical and non-medical ethicists argued about different (potential) decisional grounds, but also about the ‘morality’ of their own norm-setting powers.⁸¹⁶ Eventually a multi-disciplinary team of specialists and scholars came up with a ‘Code Black’ scenario in which years of lived experience was the final benchmark.⁸¹⁷ The plan was eventually endorsed by the Public Prosecution Office and by the Ministry of Public Health—but not after fierce debate in which the difference between medical and political decisions was differently defined by different members. On behalf of the State, the Health Care Minister objected to the scenario with the argument that such decisions should not be made by doctors but ‘left up to chance’ instead, in honor of the equality of all life and personal dignity.⁸¹⁸ This received critique from the medical field, who opposed to the objections on material grounds but also deplored this politicalization of medical decision making.⁸¹⁹ Several members of Parliament, in agreement, filed a motion to endorse the scenario. “We politicians only perform political diagnoses,” a member of parliament argued, “and it is a political choice to trust those people who are most knowledgeable and face [the pandemic situation] every day.”⁸²⁰ In the meantime, members of the triage teams that were being set up in hospitals had become nervous and expressed to want to perform their roles anonymously. Health law scholar Leegemate responded publicly in the news, explaining how this would amount to an unlawful act: patients have a right to know who makes medical decisions about them.⁸²¹

816 Marcel Verweij and Roland Pierik, ‘Het pijnlijke gesprek over ziekenhuisbedden moet juist nu gevoerd worden’, *Bij Nader Inzien* (blog), 23 March 2020, <https://bijnaderinzien.com/2020/03/23/laat-niet-aan-artsen-over-wie-een-bed-krijgt-op-de-intensive-care/>; Fleur Jongepier, ‘Opinie: Jongeren voorrang geven op ic? Het draaiboek van “code zwart” rammelt aan alle kanten’, *de Volkskrant*, 17 June 2020, sec. Opinie, <https://www.volkskrant.nl/gs-b98d9cd4>; ‘Ethici, even pas op de plaats’, *Sociale Vraagstukken Sociale Vraagstukken: Wetenschappers & professionals over maatschappelijke kwesties* (blog), 27 March 2020, <https://www.socialevraagstukken.nl/ethici-even-pas-op-de-plaats/>; Pim, ‘Protocol: geef zorgverleners en jongeren voorrang bij extreme druk op de IC - NRC’, *NRC*, 16 June 2020, <https://www.nrc.nl/nieuws/2020/06/16/protocol-geef-zorgverleners-en-jongeren-voorrang-bij-extreme-druk-op-de-ic-a4002978>.

817 IE not ‘expected quality of life’ but ‘life lived’ prevailed. Additional choices were to extend no privilege for Covid-patients over others, to extend privilege to high-intensity COVID medical carers, and include predicted IC-bed occupancy time in decisions. ‘Triage scenario non-medical considerations IC admittance in COVID-19 pandemic phase 3 (version 2.0)’ (KNMG & Federation Medical Specialists, November 2020), https://www.demedischspecialist.nl/sites/default/files/Draaiboek%20Triage%20op%20basis%20van%20niet-medische%20overwegingen%20IC-opnametvfase%203_COVID19pandemie.pdf.

818 Tamara van Ark, Minister of Health Care and Sport, ‘Kamerbrief 1801920-216248-PZO over Draaiboek Triage op basis van niet-medische overwegingen voor IC-opname ten tijde van fase 3 in de COVID-19 pandemie’ (Ministerie van Algemene Zaken, 4 January 2021).

819 ‘Kabinet wil geen leeftijdselectie op intensive care, artsen houden vast aan draaiboek’, 5 January 2021, <https://nos.nl/artikel/2363133-kabinet-wil-geen-leeftijdselectie-op-intensive-care-artsen-houden-vast-aan-draaiboek>.

820 ‘Plenair verslag Tweede Kamer, 40e vergadering’, 5 January 2021, https://www.tweedekamer.nl/kamerstukken/plenaire_verslagen/detail/2020-2021/40.

821 Johan Legemaate, ‘Triage bij “code zwart” in zorg kan niet anoniem’, *NRC*, last consulted 28 April 2021, <https://www.nrc.nl/nieuws/2021/04/26/triage-bij-code-zwart-in-zorg-kan-niet-anoniem-a4041319>.

The scenario did not have to be used in the end. If it would have, court cases were bound to follow. Self-regulation, law and policy ‘meet’ publicly in situations of conflict resolution, the outcomes of which further inform the development of both types of norm setting. Resolution happens on different levels: there are informal complaints procedures, procedures governed by disciplinary bodies, and by the public judiciary: both private (tort law) and public (criminal liability) procedures. But public court judges are generally not also medically trained. They necessarily lean on expert medical opinion to form their legal opinion about the situation at hand,⁸²² such as about what information a patient should have received. This brings in cross-disciplinary understanding challenges. A Swedish study coined a useful phrase to illustrate this problem. The authors argue that judicial governance of the medical field suffers when courts make naive use of ‘importing notions’: terms with which law refers to medical concepts that are of relevance to legal governance.⁸²³ Their example is ‘science and proven experience.’ In Swedish law this term is used to refer to medical benchmarks that law expects doctors to keep to.⁸²⁴ But since there is no medical or legal consensus, let alone a stable *shared* understanding of what the term (and its elements) means the authors argue that it cannot facilitate the meaningful medical-legal dialogue it is relied on to do. E.g., both ‘proven’ and ‘evidence’ already have different (and at times problematic) histories in both medicine and law. When such terms are used as self-evident, “medico-legal pseudo-agreements” establish. Rather than facilitate important questions and discussions, such misunderstandings obfuscate the use of powers in the field that law intends to govern.⁸²⁵

5.3.2.2 ‘Information duties’ as contractual obligations in WGBO: ambition and reception

Before the WGBO’s enactment in 1994, informed consent rights of patients could be grounded on the Dutch Constitution’s articles 10 and 11 that determine the right to self-determination and of physical integrity respectively.⁸²⁶ Although the regime ‘worked’ in the sense that explanation norm development was visible in both legal and medical practice, reports revealed that many doctors still ignored them. Physicians overestimated their own expertise, routinely manipulated their patients towards

822 Anne Ruth Mackor, ‘Rechtsregels en medische richtlijnen. Een rechtsfilosofisch perspectief op de aard en functie van regels’, in *Medische Aansprakelijkheid*, edited by S. Heirman, E.C. Huijsmans, en R. Van den Munckhof, Kenniscentrum Milieu en Gezondheid (Wolf Legal Publishers, 2016); As mentioned, these discussions are not new, see e.g. on this particular point, J Legemaate, *Goed recht: de betekenis en de gevolgen van het recht voor de praktijk van de hulpverlening*, Preadvies Vereniging voor Gezondheidsrecht (Utrecht, 1994), 46.

823 Lena Wahlberg en Johannes Persson, ‘Importing Notions in Health Law: Science and Proven Experience’, *European Journal of Health Law* 24, nr. 5 (10 November 2017): 590 emphasis mine.

824 Comparable with other references to ‘professional standards’ or a technical or practical ‘state-of-the-art.’

825 Wahlberg and Persson, ‘Importing Notions in Health Law’, 590.

826 Leenen et al, *Handboek gezondheidsrecht*, sections 1.3, 1.5.

treatment compliance,⁸²⁷ and professed to lack the time (and personnel) to inform patients about ‘highly complex medical knowledge.’⁸²⁸ The findings reflect those of Katz in the preceding part of the chapter.

A clear and urgent need for more explicit legal guidance was acknowledged, and a drafting process was started. Political perception of GP practice was of shaping influence on the legal choices that were made. Lawmakers wished to rely on (especially) GP’s to ‘deal with’ what was perceived as a very demanding, and therefore costly, patient community. Alongside the strengthening of patient rights, the protection of medical professional responsibility was made part of the drafting aim.⁸²⁹

The regime of choice became Private Law: the domain that governs the ‘horizontal’ relations of citizens, legal entities and organizations.⁸³⁰ The WGBO was designed to govern the treatment relationship as one of legal agreement (or ‘contract’) between doctor and patient. Doctors commit to perform a medical service and patients commit to meet obligations to make this performance possible.⁸³¹ A misbehaving patient may be found ‘in breach of contract’ and a doctor may end the treatment agreement on that basis, although such a breach is not easily assumed.⁸³² The legal framing of the relationship as contractual invites other, tort-related principles as well, one of which is the expectation that the contracting parties perform their obligations with ‘reasonable care.’ This raises concerns: confusion, or conflation of the different meanings of ‘care’ in medicine and tort law (i.e. between *giving care* and *taking care*) may entice doctors to avoid the provision of risky but necessary treatment, or to keep to a practice standard as a legal token of ‘care’ even when the medical beneficence of that standard is or has become unclear.⁸³³

With regard to explanation, it was considered that doctors were best placed to determine the details of patients’ information needs on a case-by-case basis. The explanation rules were therefore drafted in a very general fashion (see below). They were critically received by physicians. Among other things they argued the rules made them stage an ‘event’ around an arbitrary decisional moment, thereby interfering with the natural flow of explanation during the whole treatment process. Health law scholar Legemaate objected:

827 Legemaate, ‘The Development and Implementation of Patients’ Rights’; Leenen et al, *Handboek gezondheidsrecht*.

828 Legemaate, *Goed recht*, 58.

829 Legemaate, ‘Nieuwe Verhoudingen in de Spreekkamer: Juridische aspecten’, 9.

830 Where ‘vertical’ relations, those between citizens and state are governed by public law. Doctors who harm patients are may be (and are) prosecuted by the State for criminal medical accountability, private law procedures to retrieve financial damages typically run simultaneously.

831 One obligation for patients that is found to be especially awkward and is rarely used in practice was already named: according to the law, patients are obliged to provide all the necessary information.

832 Leenen et al, *Handboek gezondheidsrecht*, 156–57.

833 Legemaate, *Goed recht*, 62; Eric Tjong Tjin Tai, ‘Zorg, privaatrecht en publiekrecht: van ondersteuning naar handhaving, en terug’, *Recht der Werkelijkheid*, 2010, 20.

while understanding explanation as a process is certainly valuable,⁸³⁴ furthering informed consent's fundamental aims required that its core contents, as well as its pre-treatment character was pulled out of the 'implicit' sphere of medical decision making, and out from under the dominance of medical decision makers.⁸³⁵

The concern turned out to be legitimate. Around the millennium, evaluations of the WGBO revealed a gross lack of compliance with the explanation obligations. Some reports showed a *decline* of patient talking time and fewer patient questions asked in GP consultations. Doctors provided highly directive information.⁸³⁶ They left risks and alternatives undiscussed on the basis of the argument that such information was redundant (patient knowledge was assumed, risks were regarded as too small to warrant a need to know about them), or that providing such information was not conducive for treatment compliance.

The evaluations inspired intervention via a program to clarify, guide, and improve the WGBO's implementation. Practitioners, patient bodies, lawyers and health (law) scholars were engaged to improve informed consent practices. Among other things this resulted in model practice guidelines and an 'instruction/manual,' issued by the Royal Dutch Medical Association ('KNMG').⁸³⁷ The guidelines had the dual function of explaining and explicating the legal provisions on the one hand, and as an instrument of self-regulation to be referred to *by* law as a professional standard on the other. For the subject of informing and explaining, the KNMG guidelines selected GP practice as a 'standard-setting setting' to provide guidance to practitioners across the medical domain. The document places informational exchanges of, and mutual agreement between doctors and their patients central to the well-being of patients and argues that no trust relationship in fact exists when the informational relationship is of too little quality.⁸³⁸

The same issues that inspired the implementation program were found again a decade later. By that time however, SDM was on the rise in practice. It was seen to have led to some explanation progress already, and consolidating it was expected to led to more improvement still.⁸³⁹ In 2020, the WGBO's explanation rules were updated to reflect

834 Citing Stephen Wear, who argued that "no static, generic ritual can legitimately pursue the quite variable goods and values that may be at stake with different patients in different situations," Stephen Wear, *Informed Consent: Patient Autonomy and Physician Beneficence within Clinical Medicine*, Clinical Medical Ethics (Springer Netherlands, 1993).

835 Legemaate, *Goed recht*, 59.

836 'Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming', 12.

837 'Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming'.

838 'Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming', 10.

839 ZonMw, 'Achtergrondstudies zelfbeschikking in de zorg', 139–40 The guidelines oblige doctors to ensure that patients aren't influenced, nudged, pressured, or forced in their decision making. But concerns about patients' abilities to act in their best interest are noted in forceful terms, as well: a doctor who takes their patient seriously should engage "their full capacities of conviction" and if necessary persuade their patient to choose options that "apparently serve his [sic] interests." The phrasing is remarkable, but more research would need to be done to be able to properly understand how it is received.

this. Below, the new and the old obligations are presented in one text: passages that were canceled are marked as struck, and obligations that were added are printed in italics. This is done to support the later sections' discussion of how SDM came to change the law's explanation rules – too much according to some, not enough in the eyes of others. But it also makes immediately clear that at least attempts were made in law to push back on some GP 'steering': for example, 'treatment that the care provider finds necessary' is changed into 'suggested treatment' (2a), informing is now accompanied by 'engaging in timely discussion' (1) and the situation and needs of the patient are explicitly made to count. (3)

Among the things that haven't changed is the fact that nowhere in the obligations is the word 'explanation' used. The choice is surprising, at the least, in light of how little information in this domain is self-evident.

5.3.2.3 *Legal text and recent amendments*

Dutch Civil Law, Book 7, Title 7, Section 5 (WGBO), Article 7:448

1. The care provider informs the patient clearly ~~and, if requested, in writing~~ *in a way that is appropriate to his comprehension, and engages him in timely discussion* about the planned examination and suggested treatment and about developments related to the examination, the treatment, and the patients' state of health. The care provider informs patients who has not yet reached the age of twelve years in a way that is appropriate to their comprehension. 2. In complying with the obligation laid down in paragraph 1, the care provider will be guided by what the patient should reasonably know about: **a.** The nature and purpose of the *proposed* examination, the *suggested* treatment or medical procedures to carry out; ~~that the care provider finds necessary~~ **b.** The consequences and risks of the proposed examination, the suggested treatment and medical procedures to carry out, and of abstinence from treatment, with regard to the health of the patient; **c.** Other possible methods of ~~eligible~~ examination or treatment, *potentially carried out by other care providers*; **d.** The state of, and expectations for, the patient's health in relation to the field of *the possible methods of examination or treatments*; **e.** *The period within which the possible methods of examination or treatments can be carried out as well as their expected duration.* 3. *During the discussion the care provider informs himself about the situation and needs of the patient, invites the patient to ask questions, and upon request provides written or electronic information with regard to what was said in paragraph 2.* 3 4. The care provider may only withhold the above-mentioned information from the patient when providing it would clearly cause him serious harm. If the patient's interests so require, the care provider must give the information to a person other than the patient. When the risk of above-named harm no longer needs to be feared, the information shall then be provided to the patient. The care provider shall not use the authority referred to in the first sentence without having consulted another care provider on the matter.

Article 7:449

If the patient has expressed that he does not want to be informed, information shall not be provided, unless the interest of the patient is outweighed by the harm to himself or to others which may result from withholding the information.

5.3.3 Protecting patients from ‘the risk that they cannot self-determine’

5.3.3.1 Introduction

This section discusses the first aim of the WGBO’s information obligations: to establish patients’ uncoerced (‘free’) consent for diagnostics and treatment. Notwithstanding the colloquial understanding of ‘consent’ as a response to a proposition, in medicine, there is not much that a doctor should do at all without consent. It is already needed to perform examinations. The need to seek consent before entering the intimate sphere of the patient safeguards the patient’s autonomy by protecting their physical integrity and right to self-determination.

The section is structured as follows: first, two different explanation-relevant, medical-ethical views on the establishment of non-coerced, autonomous choice are introduced. Medical ethics directly inform the field’s self-governance, and are referred to more indirectly by law (see the earlier discussion on power-sharing). The arguments of the different, well-known voices pertain to the medical research trial context. This is obviously different from GP, but because the stakes are high in such contexts (false incentives to participate are a known problem, and the need for patients to trust responsibly is very high), the context brings useful concerns and differences of insight to the fore. The discussion functions as a necessary backdrop to the discussion of the WGBO’s rationales that follow. The law is put in context of the Dutch medical-legal discourse and critical perspectives in them. The section ends with a discussion of the WGBO’s address of patients’ rights *not* to know, and other exceptions the informed consent obligations: an area where its guidance is especially minimal.

5.3.3.2 Different ethical approaches of informed autonomous decision making: a very short introduction

Questions with regard to what it means for patients to make autonomous decisions, and how doctors should protect and support them in doing so are salient to the domains of medical ethics. ‘Medical ethics’ is used here as an umbrella term for a diversity of orientations and traditions. Matters around individual care and treatment are generally addressed as ‘bioethics’ and ‘medical ethics,’ where ‘health ethics’ includes broader socio-ethical, or public-oriented questions.⁸⁴⁰ Widdershoven discusses six approaches that gained traction in the Dutch medical fields in lieu of the move towards less

⁸⁴⁰ Willems and Hilhorst, *Ethische problemen in de huisartspraktijk*.

authoritarian practices; when questions of patient autonomy started to require more guidance than ‘Hippocratic virtue-ethics’ could provide. The author cites principled approaches; phenomenological, narrative, and hermeneutical approaches; discourse and care ethics approaches.⁸⁴¹ Willems and Hilhorst argue that GPs use many different types of ethical reasoning intuitively in their daily practices.⁸⁴² This section discusses two views on ‘non-coercive self-determination’ that reside on different ends of the broader ethical spectrum. As said, both pertain to consent for partaking in medical research.

The first ‘take’ is that of Faden and Beauchamp. They build on the bioethical principles of Beauchamp & Childress, which are much referred to in AI governance discourse.⁸⁴³ The authors present their development of autonomous choice⁸⁴⁴ as ‘non-normative’ for how it describes ‘objective’ conditions that allow to establish informed consent.⁸⁴⁵ A first choice they argue for is to focus on the patient, rather than the doctor, because the causal influence of intentional, ill-motivated manipulation by physicians is (too) hard to prove. The better indicator, they argue, is “substantial patient understanding.” This too is hard to establish, but for this they present a solution. What they do *not* propose to make obligatory is to verify patients’ *actual* understanding. A range of arguments are given for this: it would require a lot of time, feedback loops and recall tests; patient-side obstacles such as information overload and stress would need to be ‘overcome’; mismatches between the doctor and their patient’s class, language, and belief systems need to be dealt with.⁸⁴⁶ The authors therefore suggest to establish the *absence* of “objectively perceptible controlling influences.” Such influences do not reside in the (relevant, but inaccessible) “complex morass of human motivations,”⁸⁴⁷ nor in patients’ socioeconomic circumstances. No matter how dire the latter may be, such circumstances do not sufficiently control the will of “influencees” and are

841 Guy Widdershoven, *Ethiek in de kliniek: hedendaagse benaderingen in de gezondheidsethiek* (Boom Uitgevers, 2000).

842 Willems and Hilhorst, *Ethische problemen in de huisartspraktijk*.

843 This is also critiqued, not in the least because B&C’s ‘principled approach’ has been a weak instrument in making medicine a safer, fairer practice. The AI and Robotics group at the Tilburg Institute for Law, Technology and Society, ‘Response on the draft ethical guidelines for trustworthy AI produced by the European Commission’s High-Level Expert Group on Artificial Intelligence.’, 31 January 2019; Foster, *Human Dignity in Bioethics and Law*; Katz, *The Silent World of Doctor and Patient*.

844 The part I cite from is the ‘theoretical’ part of the book, which comes after accounts of moral and legal foundations of informed consent, and ‘descriptive’ historical developments in medicine, law, and (research) ethics. Ruth R. Faden and Tom L. Beauchamp, *A History and Theory of Informed Consent* (New York: Oxford University Press, 1986).

845 The scare-quotes are deliberate: the thesis has argued that no such thing as non-normative knowledge exists.

846 The verification of patient understanding is made obligatory in Dutch law in research contexts and some other specialized fields (e.g. organ donation), but not in the WGBO, as the next section discusses – although the KNMG model guidelines and manual do include this demand in their WGBO implementation guidelines.

847 Faden and Beauchamp, *A History and Theory of Informed Consent*, 357.

therefore of no concern to respect for autonomy.⁸⁴⁸ Faden & Beauchamp consider how dire circumstances do pose *moral* concerns about *justice*, e.g., about the just societal allocation of research risks, and about social arrangements that allow for the exploitation of persons in precarious states. But as public facing questions for research ethics, they find this to be beyond scope.

After isolating the patient from their circumstances this way, they focus on what they call the influencee's 'role constraints': states and conditions that make patients inclined to accept 'false beliefs', or more simply, anything their doctor says. Examples are the frail, the elderly, the poor, the poorly educated, the retarded, the seriously ill, the hospitalized, the institutionalized, prisoners.⁸⁴⁹ Excluded are more subtle 'aspects of identity.' Much like their dismissal of the "morass of human motivations," they warn not to get caught up in understandings of relational autonomy that undermine any belief in such a thing as 'autonomous action.'⁸⁵⁰

The need to focus on the relations of choice-making patients is prominent in O'Neill's developments: relations with who provides them with the information that their *informed* consent is based on, but also with the 'hidden' creators of such information. Because trust, she argues, is inevitably involved in 'autonomous' choice making. Since patients can never be completely informed, their trust is not (just) invested in information, but (also) in the person the information came from. And such trust extends to e.g., the designer of choosing aids that are presented to them by their doctors.⁸⁵¹ She therefore argues to improve the understanding of the different levels on which information is (and can be) provided, as levels on which trust in medicine can be established. Her suggested description of informed consent's purpose is "the reasonable assurance that patients are neither deceived nor coerced, and can judge for themselves that they aren't."⁸⁵² The aim, she argues, should be to help patients

848 Faden and Beauchamp, 358–59.

849 Their distinction between 'states and conditions' and 'circumstances' is hard to follow; e.g, financial need was for example excluded as relevant, 'poor', apparently, is something else. The definition of 'state and condition' is also unclear, and the relation of the examples to the inclination to accept false beliefs are ungrounded.

850 "[i]f one understands the entire fabric of social experience in terms of social roles and expectations—a venerable sociological tradition—role constraints might be said to operate pervasively to limit or (depending on one's point of view) altogether undermine any capacity for autonomous action .. nothing short of a full theory of the self may be required in order to resolve the many thorny ambiguities presented by such situations." Faden and Beauchamp, *A History and Theory of Informed Consent*, 268.

851 O. O'Neill, 'Some Limits of Informed Consent', *Journal of Medical Ethics* 29, nr. 1 (1 February 2003).

852 O. O'Neill; Here as in other work, she builds on the research-context rule from the Neuremberg Code: "...to exercise free power of choice, without the intervention of any element of force, fraud, deceit, duress.. or other ulterior form of constraint or coercion.." Onora O'Neill, 'Accountability, Trust and Informed Consent in Medical Practice and Research', *Clinical Medicine* 4, nr. 3 (1 May 2004): 269–76. "Some Limits of Informed Consent," *Journal of Medical Ethics* 29, no. 1 (February 1, 2003).

make ‘sound judgment calls’ to invest their trust or not, and subsequently to help them in their decision making on the basis of the trusted information. Some types of information may not perform so well seen from this understanding; or even have adverse effects on the provision of health care as a whole.

Her example of adverse effects is also named by Dutch author Trappenberg. It concerns the provision of ‘performance indicators’ of treatments, institutions, and health care providers. Rather than indicators of trustworthy practices, such performance scores may become indicators of institutions and carers who have started to avoid rare, risky, and/or complex procedures to ensure a good performance score, therewith impoverishing the provision of public health care as a whole.⁸⁵³ An illustration of ‘poorly performing’ information provision exists in a 1999 UK study on treatment choosing aids. Same as in The Netherlands, doctors and patients in upcoming SDM practices in the UK had a very large need for such materials.⁸⁵⁴ Early choosing aids were however found lacking on many points ranging from issues of graphic design, understandable and respectful (rather than patronizing and dismissive) language, to issues of unmet patient needs. Among these were a lack of contextual information (such as causes and consequences of conditions and treatment), and a lack of honesty with regard to uncertainty, quality, and availability of scientific evidence. Connecting to O’Neill’s point of extended trust, they also note the fact that the aids’ authorship was obscure as a problem. Based on their study, the authors created a list of 12 general reasons that patients in SDM need information for.⁸⁵⁵ These include the need for themselves and their social relations to understand (not just to ‘know’) what ails them: a ‘relational’ understanding of autonomy. It also includes the goal of helping patients identify further information and modes of support,⁸⁵⁶ broadening the spectrum of information to help patients decide to ‘invest their trust or not.’

853 O’Neill, ‘Some Limits of Informed Consent’; Margo Trappenburg, ‘Ik en mijn medepatiënt: Juridiseren in de gezondheidszorg’, *Recht der Werkelijkheid*, 2010, 12; In 2010, Maastricht court decided that a physicians’ lack of experience with a certain treatment can pose a risk that needs to be ‘covered’ by informed consent procedures J. Legemaate, ‘De informatierechten van de patiënt: te weinig en te veel’, *Tijdschrift voor Gezondheidsrecht* 35, nr. 6 (2011): 480.

854 A. Coulter, V. Entwistle, and D. Gilbert, ‘Sharing Decisions with Patients: Is the Information Good Enough?’, *BMJ* 318, nr. 7179 (30 January 1999): 318–22,

855 Patients need information to: • Understand what is wrong • Gain a realistic idea of prognosis • Make the most of consultations • Understand the processes and likely outcomes of possible tests and treatments • Assist in self care • Learn about available services and sources of help • Provide reassurance and help to cope • Help others understand • Legitimise seeking help and their concerns • Learn how to prevent further illness • Identify further information and self-help groups • Identify the “best” healthcare providers Coulter, Entwistle, en Gilbert.

856 The earlier named example of the breast cancer treatment choosing aid that ignored concerns about labour incapacity clearly failed on this point. Engelberts, Schermer, and Prins, ‘Een goed gesprek is de beste persoonsgerichte zorg’.

5.3.3.3 *Patient, person, contractual partner: choices in law about whose ‘free choice’ to serve*

Dutch health law scholars have voiced concerns about unhelpful understandings of autonomous patient choice that they see expressed in the WGBO, and in Dutch Health Law more broadly. The law’s strong ‘freedom from’ perception on patient self-determination is criticized for how it idealizes the individual, capable, autonomous citizen-patient. A frame that is not protective enough of what in reality are persons in vulnerable states, who also enter the ‘contractual relationship’ on a considerable informational disadvantage.⁸⁵⁷ The Law’s emphasis on self-determination is seen to take attention away from Health Law’s two other ‘main principles:’ protection and equality.⁸⁵⁸

The concerns are supported by different WGBO evaluations. The first studies had revealed that GPs’ patients had not become more active participants in consultations. As was cited earlier, they asked fewer questions and less ‘talking time’ was spent with them in general. The findings were interpreted in different ways. Some writers were concerned that patients did not ask questions when they (falsely) assumed they had already been told everything of importance *because* the informed consent obligations had been made part of law.⁸⁵⁹ Although the concern seems to assume that patients ‘know the law,’ the argument can be read as a case to verify what patients actually understand more actively. Others suggested that patients were simply not (yet) as ‘autonomous’ as the WGBO expected them to be: that they attached more weight to their GP’s opinions than the law assumed they would.⁸⁶⁰ They argued that SDM and patient-centered communication methods (on the rise in GP practice at the time) might be ‘better placed’ *than* law to channel patients’ needs and wishes.⁸⁶¹ The argument seems to assume that these developments could perform well enough *without* further codification, but that trust was not well placed. A legal background study from 2013 suggested that these patient-centered methods were themselves being overshadowed

857 J.G. Sijmons, ‘De stimulerende middelen van de wetgever (2007)’, in *Oratiebundel Gezondheidsrecht: verzamelde redes 1971-2011* (Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012).

858 Aart Hendriks, ‘In Beginsel (2005)’, in *Oratiebundel Gezondheidsrecht: verzamelde redes 1971-2011* (Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012), 380; Sijmons, ‘De stimulerende middelen van de wetgever (2007)’; See also Hendriks who adds how the Law’s idealizations of autonomous choice ignore the reality that patients’ choices are in fact quite limited since much is predetermined in public health policy, insurance schemes, et cetera – this was discussed in the preceding part of the chapter Aart Hendriks, ‘Challenges and Obstacles to Access to Justice in Health Care’, *Recht Der Werkelijkheid* 36, nr. 3 (November 2015): 127–38.

859 W. R. Kastelein, ‘Patiëntenwetgeving: Bureaucratie of Bescherming? (2001)’, in *Oratiebundel Gezondheidsrecht: verzamelde redes 1971-2011* (Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012).

860 Roland Friele and Remco Coppen, “Wetgeving en de positie van de patiënt: instrument voor verandering of terugvaloptie?,” *Recht der Werkelijkheid*, 2010, 14.

861 Roland Friele and Remco Coppen.

by the uptake of EBM and (other) protocolization trends.⁸⁶² Especially for GP practice, this was understood to pose risks to the quality of doctor-patient engagements.⁸⁶³

The 2020 edition of the Handbook of Health Law treats the WGBO as it was updated after these evaluations and several background studies. The Handbook writes that the legal informed consent paradigm means to protect patients ‘from the risk that they cannot self-determine,’ not from medical risk itself.⁸⁶⁴ The statement is understandable in light of the ‘inherent risk’ of medicine, but in how the juxtaposition makes the provision of information instrumental to self-determination, it arguably downplays the value of informing patients regardless of their options. This can be criticized in light of the principle of ‘equality’ and the large information inequality between the parties, and in light of the earlier sections’ discussion of the role of informational exchanges for the establishment of doctor-patient trust that was described in the health professionals’ WGBO implementation guide.

With regard to what patients ‘reasonably’ need to know to self-determine (article 7:448/2, printed above) the Handbook cites how the following general norm for guidance was established through case law: “the provision of information about the facts and possibilities that a reasonable person could be expected to consider before making a decision, or that he [sic] needs to inform further behavior.”⁸⁶⁵ In how the word explanation is not used, and person is chosen over patient, the norm (again) seems to express its contract-law embedding: the parties are treated as theoretical equals.

Examples of information that needs to be provided include expected pain/recuperation time; reasons for, and contents of, diagnostic and treatment referrals; and ‘frequent or important’ risks and side effects of proposed treatment. The Handbook states that very rare risks generally don’t need to be communicated. At least, not in precise terms, unless the consequences upon their materialization are very large.⁸⁶⁶ The logic here according to Legemaate is that patients shouldn’t be ‘needlessly scared.’⁸⁶⁷ And he is critical of how the courts (both disciplinary and public) drew such influential normative conclusions since these are in contradiction with (empirical) research that showed different patient preferences.⁸⁶⁸ More on the challenges of explaining risks is discussed in a later section; the point to dwell on here is that if patients want to

862 ZonMw, ‘Achtergrondstudies zelfbeschikking in de zorg’.

863 Section 5.2.3.2

864 Leenen et al, *Handboek gezondheidsrecht*, 654.

865 Leenen et al, 118.

866 The example used is the risk that a patient’s womb and ovaries need to be removed. Leenen et al, 119.

867 Legemaate, ‘Patiëntveiligheid en patiëntenrechten (2006)’, 290.

868 The courts in general seemed to side with carers in what Legemaate calls the ‘preference paradox’ between doctors and patients: patients generally want more than doctors think they do. Johan Legemaate, “Patiëntveiligheid En Patiëntenrechten (2006),” in *Oratiebundel Gezondheidsrecht: Verzamelde Redes 1971-2011* (Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012), 290.

prove that their consent was invalid, the current legal paradigm requires that they ‘convincingly argue’ that they would have made a different choice if they had been differently informed.

Arguably, this again raises questions about Dutch law’s ambitions with regard to promoting a more equal information relationship. Established insights that ‘medical honesty’ helps patients deal with (real and possible) adverse effects of treatment that they’ve already chosen to undergo are for example ignored.⁸⁶⁹ The handbook does name some rare cases in which the right to ‘simply’ know was closer to being established. It tells of a woman who awakened from a hip operation with two legs of different length. This was unavoidable, but the woman contended that had she known, she would have opted for an operation on both hips instead of one, where instead she now had to undergo a more risky recovery operation. The court of appeals established a breach of autonomy and physical integrity (a decision that was overruled by the Supreme Court).⁸⁷⁰ Another case involved a foreign language patient who was taken by surprise by life-threatening side effects of her necessary medication. She had not been told, nor could she have gleaned this information herself from the Dutch medication documentation.⁸⁷¹ The Handbook writes that a more extensive duty to inform patients does exist in cases of *unnecessary* medical interventions. The argument is that “good and proper information” is more important when “the free choice of patients is really at stake.”⁸⁷² The logic here is that people need to be protected from involuntarily harming themselves in absence of a medical need to do so. Such logic however ignores the voluntariness of undergoing ‘necessary’ harmful treatment, and pays little regard to the role of informing well for the beneficence of the treatment relationship.⁸⁷³

5.3.3.4 *Exceptions to the legal duty to inform: choices and guidance in the WGBO*

The focus of the WGBO on the right make informed choices *for* treatment also ignores the need for GPs to understand their explanation obligations in cases of ‘informed dissent,’ the informed rejection of (further) diagnostics or treatment. “There is an asymmetry here,” Willems and Hilhorst write, also with regard to the lack of research on the type of reasoning of patients in such cases.⁸⁷⁴ Pushing treatment is obviously out of the question, but so, arguably, is pushing information. The voluntariness to ‘undergo medicine’ also stretches to patients’ ambitions to know about it: they do not have an unqualified *duty* to be informed.⁸⁷⁵

869 Legemaate, *Goed recht*.

870 Leenen et al, *Handboek gezondheidsrecht*, 655 Hof Arnhem 4 December 2007, ECLI:NL:GHARN:2007:BM5197 and HR23April 2010, JA2010/97.

871 Leenen et al, 655.

872 Cases on cosmetic surgery and artificial insemination are named Leenen et al, 119.

873 ‘Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming’, 10.

874 Willems and Hilhorst, *Ethische problemen in de huisartspraktijk*, 33.

875 Leenen et al, *Handboek gezondheidsrecht*, 126.

Law's current guidance here is minimal. Several insights from the ethical domain could possibly inform a more ambitious agenda. For example, imperatives of informing are also grounded on the notion that beneficial patient trust improves when patients know that their GP's are willing (and able) to talk about painful subjects.⁸⁷⁶ And so, when a GP is confronted with a patient unwilling to discuss the lethality of their disease, they will need to find out what that wish represents: whether the patient does not want to know (which needs to be respected), or that something else is at stake. They could be scared of what happens next (including how they are able to deal with knowing), or scared or uncomfortable to discuss the subject in their social circle.⁸⁷⁷ They may fear that the process of dying is sped up by their awareness of the verdict.⁸⁷⁸

Doubts about the beneficence of medical honesty also arise in relation to sharing suspicions of disease. It is established that diagnostic knowledge can affect patients' emotional as well as biological states and that the effects of a 'fake news' diagnosis can persist even after the wrong is cleared up.⁸⁷⁹ Ethical-legal guidance on this subject holds that if waiting for (more) certainty can be done, burdening patients with diagnostic assumptions may breach a patients' moral as well as legal right to not be informed.⁸⁸⁰

In all cases, GPs will need to balance their patients' wish not to be informed against possible adverse consequences for patients themselves and others. This much is indeed acknowledged in law. Some highly impactful diseases are for example very treatable or even curable, or became more treatable over time, such as HIV infection. A patient's wish to remain ignorant of impactful diseases may be reconsidered in such cases. An obligation to inform may even resurrect when patients' (expected) states render them dangerous to others; e.g. when their capacity to drive is affected.⁸⁸¹

Another point on which the Handbook is explicit is the use of placebos. When a placebo is used, the legal duty to inform can be suspended to protect the therapeutic goal, and the existence of a therapeutic goal means the prescription of a placebo is "not paternalism."⁸⁸² It is not explained how this view is to be reconciled with the general therapeutic goal of medical honesty and (more) information equality.⁸⁸³ Katz, for example, argued for honesty in service of responsible patient trust-investments: they need a realistic understanding of the medical sciences. The Handbook's standpoint arguably also does little justice to the complexity for decision making that follows

876 Willems and Hilhorst, *Ethische problemen in de huisartspraktijk*, 29.

877 D. Willems and M. Hilhorst, *Ethische Problemen in de Huisartspraktijk*, Practicum Huisartsgeneeskunde (Bohn Stafleu van Loghum, 2016), 29.

878 Willems and Hilhorst, *Ethische problemen in de huisartspraktijk*, 28.

879 Willems and Hilhorst, 30.

880 Willems and Hilhorst, 30.

881 Leenen et al, *Handboek gezondheidsrecht*, 126.

882 Leenen et al, 124.

883 Legemaate, *Goed recht*, 60.

from the ever-evolving research on placebic (and nocebic) effects, saliently at play in GP practice with regard to for example certain categories of anti-depressants.⁸⁸⁴ More fundamentally, the argument sits in tension with the law's own conditions to legally invoke the 'therapeutic exception' to the duty to inform. Article 7:448/4 says information may be withheld only when severe harm can be expected to arise from informing.⁸⁸⁵ Constituting documents and legal evaluations both emphasize that such a situation will rarely exist. E.g., the risk that an informed patient may be scared out of treatment does *not* count, nor does the argument that a state of blissful ignorance would benefit their healing process. An "objective and verifiable" risk must exist, confirmed by a second physician.⁸⁸⁶

5.3.4 Supporting patients' decisional capabilities: the legal uptake of SDM

5.3.4.1 Introduction

This section builds on the discussion of SDM in the previous part of the chapter by discussing the legal uptake of this practice concept with regard to explanation. The section therewith tends to the 'positive freedom' side of free, informed consent, focusing on the ambition of the WGBO's obligations with regard to ensuring that GPs enable their patients to participate responsibly in what was argued to be a knowledge making, and not just decision making practice.

The chapter first discusses the SDM-inspired amendments to the WGBO. It places them in a critical perspective of the earlier SDM discussion, this time adding legal arguments and considerations. It then discusses two subjects in which the social-epistemic inequality of the GP-patient relationship elicits useful thinking about the role of law: complaints procedures and discussions about medical risks. This completes the second part of the chapter.

5.3.4.2 Added SDM obligations, still a weak promotion of the 'right' informed consent?

As Legemaate argued, 'perpetual care' is required to substantiate legal aims in the high trust, and highly personal, medical practice. Citing one of Dutch Health Law's founders Leenen, Legemaate describes law as an instrument in service of an ideal, where "[I]aw

884 See for example Laura de Wit et al, 'Antidepressiva in de dagelijkse praktijk', *Huisarts & Wetenschap* 2019, nr. 12, last consulted 19 November 2022; A very recent book engages with how Western medicine is seen to struggle with the phenomenon Kathryn T. Hall, *Placebos* (MIT press, 2022).

885 Leenen et al, *Handboek gezondheidsrecht*, 126.

886 The rules also state that "another person [than the patient] may be informed if the patient's well being demands it." But that too should be applied 'restrictively' in light of the additional breach of confidentiality that this would constitute, and (adds The Handbook) because such an act might backfire: the behaviour of the better informed 'others' may reveal that something's up, causing anxiety for the 'ignorant' patient. Leenen et al, 125.

is engaged to realize the best possible health care.”⁸⁸⁷ In response to physicians’ critique on (the plan to) add legal informed consent obligations, he argued how the WGBO’s SDM-inspired amendments *updated* the law’s expression of what it already intended to ‘produce’ qua informational exchanges between doctors and patients. Put differently, physicians could have complied with this broader understanding themselves, if more legal requirements was not what they wanted.⁸⁸⁸

The obligations that were added were already part of the fields’ professional standards, such as the need to discuss patients’ preferences, convictions, and circumstances. In an article on the proposed 2020 text of the informational obligations, Legemaate argues that the amendments strengthen the law’s focus on care professional’s duties, attitudes and professional behavior, explicating *how* they should serve their patient’s rights. Added (see the text of the article above) are the obligation to *discuss* the provided information, to do this timely, to cater to *all* individual patients’ understanding needs, and to invite patients to ask questions. The Handbook writes how more is now legally expected of doctors with regard to engaging with patients’ social/medical situatedness: to probe more than they did in order to elicit patient preferences, and to spontaneously inform patients of (strong) suspicion of (severe) disease when they know a patient is about to make consequential life choices (the example named is when a patient is about to take on “considerable financial burdens.”)

Notably *missing* is the obligation to verify patients’ understanding, even though this is part of the KNMG model guidelines and instruction/manual. Legemaate argues how this is unfortunate for several reasons: first of all in light of well-known understanding challenges of patients,⁸⁸⁹ and secondly because a legal obligation to verify patient understanding would allow to “check the checker”: an argument to make more use of law as an instrument to improve a practice that was found lacking.⁸⁹⁰ Legemaate argues the omission signifies the law’s still weak promotion of what SDM is, and requires, in practice. He compares the WGBO’s obligation that doctors inform themselves about their patient’s situation and needs, with the obligation to practice preference sensitive care in US president Obama’s Affordable Care Act. The ACA’s wording, he argues, more clearly expresses that the patients’ knowledge and views need to be *engaged* with. The Handbook on Dutch Health Law does acknowledge how “at least 85 percent of information” that is needed to diagnose a patient, “comes from them.”⁸⁹¹

887 Legemaate, ‘Rechtstekorten in het gezondheidsrecht’, 193–94.

888 Legemaate, 202.

889 Dink A. Legemate and Johan Legemaate, ‘Het preoperatief informed consent’, *Nederlands Tijdschrift voor Geneeskunde* 2010, nr. 154:A2492 (2010): 2.

890 Legemaate, ‘Aanpassingen van de WGBO’.

891 Leenen et al, *Handboek gezondheidsrecht*, 115.

There is much more SDM relevant guidance in professional standards that wasn't engaged with in the updated WGBO. Future evaluations will need to reveal to what extent the guidance from these domains, perhaps need legal amplification, too. Various challenges to the establishment of SDM' ideals of social-epistemic equality were named in section 5.2.3.3's SDM discussion (unhelpful power dynamics between more and less vulnerable parties; naive assumptions about patients' participatory inclinations and assertiveness; and the need for doctors as well as patients to (further) develop a range of capabilities to support their mutual, responsible understanding.)

Examples of relevant professional norm setting include the need to understand the (possible) mattering of group-aspects & societal factors (KNMG guidelines),⁸⁹² and the need to analyze a GP practice's patient population for characteristic informational needs (GP education).⁸⁹³ GP's are to show respect, take complaints seriously, and to 'know their own cognitions.'⁸⁹⁴ Cited earlier were how the KNMG guidelines qualify the informational exchanges of doctors and patients instrumental to the establishment, rather than just the quality, of a therapeutical trust relationship: informing a patient *well* is defined as an act of 'respect' that allows patients to trust their physician.⁸⁹⁵ Since patients are widely reported to have trouble recalling the content of medical conversations, especially impactful ones where this is of more importance, the KNMG encourages doctors to encourage their patients to record diagnostic conversations—but quite a number of doctors are reported to be uncomfortable for various reasons.⁸⁹⁶ The need to verify patients' understanding is included in these guidelines⁸⁹⁷ and in GP education materials: referred to as part of the constructive dialogue, verification of understanding is instrumental to the trusting working relationship's perpetual development.⁸⁹⁸ This brings in 'explanation' where the WGBO does not, and acknowledges the need for both parties to 'keep at it' in the understanding that relationships need maintenance. All examples also contrast with the earlier cited

892 E.g. personal, health incident-related, socio-economic, 'ethnic' aspects.. the lists are endless: the 2004 Model Guidelines advise to cater to the 'individual needs, priorities, and to their capacity', where capacity here means the (max.) load of information they can handle. 'Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming', 57.

893 https://www.huisartsopleiding.nl/images/837275_Themas_en_KBA_128x190_Brochure_STAND.pdf, theme 10 'organization'.

894 Characteristic professional activities Theme one: short episodes Huisartsopleiding Nederland, *Themas en KBA's*.

895 'Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming', 10, 27, 59.

896 'Opnemen van het gesprek'; Héman, 'Niet stiekem'; 'Opnemen van gesprekken door patienten: Uitkomsten raadpleging KNMG Artsenpanel'.

897 In the FAQ section, the question whether a health professional should check patient understanding is answered with 'yes', and comes qualified with a discussion of how a patient's 'type' may still make this a hard thing to do, and tips to deal with this. 'Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming', 46.

898 Huisartsopleiding Nederland, *Competentieprofiel van de Huisarts.Pdf*, 2016, https://www.huisartsopleiding.nl/images/opleiding/Competentieprofiel_van_de_huisarts_2016.pdf.

considerations of Faden and Beauchamp; which is of interest to note in light of developing plans for Dutch AI governance.

5.3.4.3 *How social inequality exacerbates informational inequality: problems with complaints procedures*

The Handbook discusses some doctor-patient relational dynamics in the context of complaints procedures. It writes how there is an understanding that informal complaints procedures are a better fit for “relational aspects of care (communication and treatment)”⁸⁹⁹ as opposed to disciplinary and public court procedures that are best reserved for “treatment technicalities.” It accedes how the two types of complaints are however hard to separate in practice, and issues brought forward as the one may in fact be about the other. This is to be expected, the authors write: patients lack the medical knowledge to understand whether they *should* have been told about something, let alone whether that omission of information amounts to a ground for complaint.⁹⁰⁰ The point makes sense, but leaves unexplored how power dynamics exacerbate this information inequality, and what kind of procedure makes sense in light of it.

Policy of GP organizations states that a complaint trajectory starts within the GP-patient relationship.⁹⁰¹ When the complaint is unresolved, it is escalated to a designated complaints person, but still for ‘mediation’: the aim of the complaint procedure is reconciliation rather than (just) resolution.⁹⁰² The first report of the recently instated Independent Complaints Commissioner raises questions about this system: many patients are reported to be reluctant or scared to address issues with their GP’s directly. Among other things, GPs were found to respond defensively.⁹⁰³ Only half of the patients that did take this route were satisfied with the treatment of their complaint, as opposed to patients who were able to file their initial complaints with a complaints officer.⁹⁰⁴

An interesting finding of the report is that patients who did complain to their GPs directly also wanted to improve their GP’s practices for other patients. To this end complainants were eager to be informed about the uptake of their complaint, and

899 Such a preference is hardly surprising, and not treated as of special interest here. Leenen et al, *Handboek gezondheidsrecht*, 610, 612.

900 Leenen et al, 632.

901 In line with the spirit of contract law, when differences of opinion arise about what information an ‘average’ person could have ‘reasonably’ expected, resolution starts with a patient filing a complaint according to the procedure that is offered by their care provider. Leenen et al, 608.

902 R D Friele et al, ‘Wet kwaliteit, klachten en geschillen zorg’, 276.

903 Emiel Stobbe, ‘Ervaringen na een klacht over de huisarts. Veertien diepte-interviews met patiënten’, *Huisarts & Wetenschap* 2020, nr. 63 (2020): 2.

904 ‘Jaarverslag tuchtklachtfunctionarissen 2019’, rapport (Ministry of Public Health, Well-Being and Sports, 30 June 2020).

disappointed as this was rarely shared.⁹⁰⁵ The procedure does not require that official statements are made upon conclusions unless the complaint is made in writing, in which case the new complaints law requires a written conclusion about the complaint within six weeks.⁹⁰⁶ Arguably, a useful opportunity is missed to learn from what happens in these resolutions, especially in light of the mutual understanding challenges reported about the field.

5.3.4.4 *Discussing risk: when making conversation turns into making preferences*

The updated WGBO obliges doctors to ‘inform themselves’ of patients’ needs. But medical reality teaches how patients’ needs, values, and preferences also establish through consultation itself.⁹⁰⁷ Patients do enter with personal notions and values, but these can only usefully inform their medical choices when they understand how their lives might be impacted by their states, diagnostics, or treatment.⁹⁰⁸ This puts weight on consultations around medical situations that require patients to consider and therefore understand risks. Even more so with regard to novel types of risk that doctors and patients are both still inexperienced to think about. Such situations are increasing as a consequence of advances in genetic testing and other ‘predictive’ diagnostics.⁹⁰⁹ The need for guidance for patients around the merits of such testing and how to deal with *possibly* life-changing information is growing. Research done on *how* to communicate risks, e.g., how to deal with patients’ cognitive grasp of quantified risk statistics is not necessarily of sufficient guidance; the Handbook advises how dealing with this requires further development of the current professional standard, for example through guidelines.⁹¹⁰

The WGBO’s uptake can be understood as a ‘careful’ legal approach to not interfere in such ethically sensitive matters; or a choice to side with doctors’ own views on matters of risk communication, which would be more in line with current Dutch Health Law tradition.⁹¹¹ As was discussed, this tradition is *not* in line with research findings

905 Stobbe, ‘Ervaringen na een klacht over de huisarts. Veertien diepte-interviews met patiënten’; Earlier research showed that complaints officials did not regard the improvement of public health to be part of their duties. Friele et al, ‘Wet kwaliteit, klachten en geschillen zorg’, 166.

906 Friele et al, ‘Wet kwaliteit, klachten en geschillen zorg’, 104.

907 Katz: “a better appreciation and, in turn, a better management of uncertainty will not emerge out of more refined technical medical knowledge, but rather out of the physician’s and patient’s psyches where, after all, certainty and uncertainty are perceived, judged, evaluated, and prepared for expression.” Katz, *The Silent World of Doctor and Patient*. 1984, 206.

908 ‘Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming’, 38.

909 Willems and Hilhorst, *Ethische problemen in de huisartspraktijk*, 15–17, 19.

910 Leenen et al, *Handboek gezondheidsrecht*, 322.

911 See previous section.

that patients typically want to know more than is discussed with them.⁹¹² It seems that doctors tend to ‘err on the side of caution,’⁹¹³ but an empirical study also mentions other reasons of doctors for not wanting to mentioning risks. Among them were their *own* difficulty in grasping risks, and concerns that when they overwhelm their patients’ cognition, their consent would not be grounded.⁹¹⁴

5.4 The WGBO’s explanation governance in terms of the modeled duties of explanation care

5.4.1 Introduction

A discussion of diagnosis as typical ‘decisions’ made in GP practice, and as a type of knowledge making that is also political started off the ‘what, who, and how’ part of this chapter’s domain study. What followed was a characterization of GPs as knowledge and decision makers, and a discussion of two decision making paradigms that are significant for ‘how’ decisions are reached. The second part of the chapter discussed the WGBO’s basic explanation obligations, placing them in the larger governance field and critical discussion.

This third part reflects on these findings through an epistemic in/justice informed lens. It relates the earlier sections to the modeled duties of explanation care. To recap, the Model categorized fundamental explanation values that require institutional expression by making them part of (any) decisional domains’ legal explanation rules, or so this thesis argues. The structure of the preceding chapter is repeated: the element is treated in consecutive sections. The element’s descriptions start off each section, followed by perceived recognition for the element’s components and an analysis of the address (or lack thereof) in the domain’s legal explanation rules.

The critical description of the domain’s legal explanation paradigm that is constructed this way serves to answer part of the third research question. The findings are picked up again in chapter 6. That chapter draws lessons from both domains’ analyses, to inform the (further) development of ruled explanation paradigms in AI-informed times.

912 Age, coping styles, information eagerness, risk-averseness, SDM-inclinedness N B A T Janssen et al., ‘Under What Conditions Do Patients Want to Be Informed about Their Risk of a Complication? A Vignette Study,’ *Journal of Medical Ethics* 35, no. 5 (May 1, 2009): 276–82.

913 “Good care relationships imply a prudent approach to the disclosure of risks.” N B A T Janssen et al, ‘Under What Conditions Do Patients Want to Be Informed about Their Risk of a Complication? A Vignette Study’, *Journal of Medical Ethics* 35, nr. 5 (1 May 2009): 281; See also Palmboom et al who advise to err on the side of risk disclosure G. G. Palmboom et al, ‘Doctor’s Views on Disclosing or Withholding Information on Low Risks of Complication’, *Journal of Medical Ethics* 33, nr. 2 (1 February 2007): 67–70

914 Palmboom et al, ‘Doctor’s Views on Disclosing or Withholding Information on Low Risks of Complication’.

To come to this part's analysis, findings of the preceding sections were again hand-coded for Model element relevance. In this domain, the relatively recent shift from purely authoritarian ('explanation-free') decision making, to the acknowledgment that patients have a right to know, and eventually, to participate in decision making was of special interest. First of all, clear support for what the Model promotes expresses in literature about this shift, and which challenges it still faces. Secondly, the shift reveals usable insights about law's dealings with the governance of an expert knowledge practice.

5.4.2 Element one: investigating explainer authority

Explainers are obliged to investigate their own social-epistemic positions with regard to their decision-making modalities, and their domain's underlying (input) knowledges in order to assess their role (=explainer) authority: does the explainers' understanding justify their authoritative and trustworthy explainer position? If no (or can't investigate), rebel.

This element obliges that explainers avoid to become an instrument of unjust ('bad', oppressive) knowledge practices, and are able to explain their 'avoidance strategies' to their explainees. To what extent they need to in fact explain these strategies is best determined in a decision domain's context. More positively expressed, this element promotes that explainers are able to communicate how, and not just that they are trustworthy 'knowledge practitioners,' and not just accountable decision makers. The point at this stage is to link the self-reflection of explainers to their position of authority *vis-à-vis* explainees, as part of responsible practice. The need for explainers to rebel exists when explainers feel incapable to do this, for example because they don't have access to justificatory sources or aren't afforded the time, or means, to investigate.

5.4.2.1 Recognition of element one: problems in the absence of non-binding obligations for self-reflection

The first element relates investigative duties to explainer roles: in this domain, those of medical experts in highly personal practices. Literature on the studied domain offered a considerable amount of clues about the merits of this linked approach. A first cohort of clues, sourced from descriptions of the pre-informed consent, pre-SDM paradigms is discussed below. The subsequent section treats sources of recognition in contemporary medical-ethical-professional literature. Finally, the take-up and address of these insights in the WGBO's explanation paradigm are considered.

General findings about the broader medical field attest to how doctor's own, responsible understanding suffered during the time that moral, ethical, and legal obligations to justify medical knowledge were absent. Katz and others described a culture of silence in which a strongly authoritarian structure between medical practitioners themselves already functioned as an obstacle to the critical progress of medical practices. The

lack of obligations to engage in explanatory, justificatory and ‘knowledge making’ conversations with patients further prevented this progress: it allowed physicians to understand their knowledge and decisions as sufficiently informed whereas they lacked insight into their patients’ needs, experiences, preferences, and understanding.⁹¹⁵

The lack of explanation obligations also allowed the cultivation of an authoritarian doctor-patient relationship as a thing of beneficence.⁹¹⁶ This prevented the development of a more responsible narrative of why doctors are, or when they should be, trusted by their patients. When critique swelled and (at least) the provision of information to prevent coercion was argued for, doctors were cited to contend that patients can’t but trust their doctors blindly since they can’t meaningfully understand the complexities of medical knowledge. Katz was cited to argue in turn that this narrative supported a plethora of unfounded and problematic assumptions: that doctors do not have other reasons for not-explaining (such as blind loyalty to their elders; lack of own understanding; paternalistic convictions, and a failure to recognize the humanity of certain groups of patients); that ‘medical knowledge’ is a clearly describable thing at all; that patients’ meaningful understanding depends on their individual knowledge of medical technicalities (whereas this was never investigated); that patients’ behavior signifies trust (rather than fear or submission) in the first place, and that *when* doctors are trusted, this is because of their knowledge rather than their social status. One especially problematic assumption supported by the ‘blind trust’ narrative, criticized by various authors, is the notion that doctors can engage in responsible investigations and decision making *without* their patients’ informed, and therewith meaningful, participation.⁹¹⁷

The subject of meaningful participation is dealt with further under the other elements; this section will now discuss several Dutch, GP-specific findings from around the shift to the informed consent paradigm.

The chapter discussed how in the first post-war decades, GPs positioned themselves as ‘patient specialists’ with a keen eye for societal factors of their patients’ well-being. With this, they distinguished themselves from the growing group of ‘disease specialists’ who were more strictly scientifically (according to the sciences of the time) and physiologically focused. Making use of their very personal positions in their patients’ lives, GPs engaged with social and societal causes as well as behavioral and psychosomatic dimensions of their patients’ ailments, which was trending at the time. In line with Katz, the beneficial influence of what became a highly paternalistic stance during this period was widely assumed: GPs came to see themselves as the ‘embodiment of good care.’⁹¹⁸

915 Sections 5.2.2.1, 5.2.3.3

916 Section 5.2.2.1

917 Section 5.2.3.3

918 Section 5.2.2.1

The shift to more scientifically grounded, as well as more honest GP practices was inspired by different developments at the same time. It fitted with the anti-authoritarian spirit of the time, and importantly came from within GP's own circles. There were growing concerns about the credibility of their experimental (social, behavioral) knowledge making, and about how the lack of knowledge about their states was not just 'beneficial' but produced anxiety in patients.⁹¹⁹

But decades of paternalist GP practice, and a lack of justificatory traditions to learn from had made both doctors and patients ill prepared for more equal interactions. The experienced plight of GPs to educate their patients in their new, autonomous roles was cited as 'new paternalism.' A different concern of GPs, especially relevant to element one was how the shift of focus to patients' self-reported problems could come at the cost of understanding systemic societal causes and factors that their specialty had made them sensitive to. They warned how treating 'societal ailments' as individual medical issues could make them complicit in consolidating problems that need to be addressed at the political level.⁹²⁰ The concern makes sense, as did the experience that patients were indeed not ready to participate on their own behalf very well. But neither were GPs, and notwithstanding increasingly elaborate professional guidelines, their practices remained highly non-compliant for many decades.⁹²¹ In that light and those of the guidance that expresses in the voiced concerns, the choice of lawmakers for a contract-law construction expresses a rather weak expression of element one's values. This will be further discussed after an exploration of the guidance that did establish in professional standards.

5.4.2.2 Promotion of critical and justifiable practices: recognition in the non-legal fields

Contemporary professional self-definitions describe GPs as continuous, multi-disciplinary learners; critical appraisers of, and contributors to, the medical sciences on the one hand, and to public health law & policy on the other. Patients are acknowledged as socially and societally embedded subjects.⁹²² Several findings from the sourced literature support these definitions.

For example, in the 'how' section, EBM's quantitative, standardizing influence was met with criticism that reflects element one's values. Although the goal of (more) scientifically justified decision making was endorsed, EBM's narrow (quantitative) understanding of science and justifiability was much critiqued, as was the idea that science naturally promotes fairer medical practices. EBM-inspired guidelines, standardization tools and decision aids raised concerns for their lack of transparency

919 Section 5.2.2.1

920 Section 5.2.2.1

921 Section 5.3.2.2

922 Sections 5.2.2.2, 5.2.2.3

with regard to authorship and embedded norms and values, and for how they did not promote the type of qualitative capabilities that the GP practice especially requires. Best-of-both-worlds kind of practices were promoted (‘Real’ EBM and ‘Evidence Informed Practice.’)⁹²³

Dutch GP education materials emphasized the need for GPs to engage with their intuition, background, and awareness, and especially warn to be alert for prejudicial tendencies in cases of unexplained symptoms. Part of the instruction for such cases is for GPs to remain upfront about their medical uncertainty *vis-à-vis* patients while they work *with* them towards a *shared* understanding of the best path forward, to remain alert for any emergent clues of somatic causes, and take their patients’ own intuitions seriously, e.g. with regard to preferred specialist referrals.⁹²⁴ The instruction is also of relevance for element three, which is focused on interaction—the reason to name it here is that it expresses the beneficence of linking accuracy & sincerity in the (pre-) diagnostic phase.

The need to investigate, and (be able to) to testify to the (possibly problematic) social-political dimension of diagnostic concepts⁹²⁵ with patients was less explicit in the studied professional guidance materials.⁹²⁶ Recognition for related concerns such as the social effects of diagnosis was, e.g., in instructions to be careful about sharing diagnostic assumptions. GPs were advised to be prudent about this for various reasons, such as the instability of diagnostic concepts, and because the act of diagnostic labeling has real-life effects on patients’ well-being that did could persist regardless of the eventual ‘verdict.’⁹²⁷ The finding resonates with the internalized uptake of wrongful knowledge claims and spheres by victims of such injustices, and is especially important in light of the many wrongful medical claims that are historically produced.

5.4.2.3 *Promotion of justifiable knowledge authority: law as the least ambitious norm setter?*

The expression of element one’s values in the WGBO’s explanation paradigm is minimal. The need for doctors to assess the fairness of their knowledge practices and (be able to) prove as much to patients is entirely absent. The Handbook of Dutch Health Law did write how doctors needed more guidance in discussing fundamental questions of medical knowledge with regard to *novel* types of knowledge to deal with, especially predictive diagnostics that come with novel types of risk considerations—

⁹²³ Section 5.3.2.2

⁹²⁴ Sections 5.2.2.3, 5.2.3.2

⁹²⁵ Section 5.2.1.2

⁹²⁶ In light of e.g., the ADHD example, such attention seems warranted – and might well exist, since no exhaustive study of professional materials was done.

⁹²⁷ Section 5.3.3.4

new for doctors too, and eliciting all kinds of value questions and ethical dilemmas.⁹²⁸ The handbook advised to further develop the professional standard to deal with this. It therewith gives no strong instrument to doctors *vis-à-vis* the new knowledge makers. GPs are explainees themselves, too. The inscrutability, also with regard to their fairness, of novel medical knowledges challenge their own understanding, with the potential to frustrate their explainer responsibility.

The advice also seems to rest on the premise that the current professional standard provides good enough guidance for the treatment of *existing* fundamental (value, ethics) questions of medical knowledge. Without a clear legal anchor for this need, it will however be hard to check compliance. The WGBO therewith expresses a weak institutional promotion of values that the thesis has argued to need strong public endorsement. In a field whose practice spheres and cultures are historically explanation and justification-averse, and in which power and information inequalities are large, that choice begs to be defended. The WGBO's lack of use of the term 'explanation' is interesting to note in this respect. Items are listed that patients need to be informed about, rather than explained, and nothing needs to be 'justified.'

Legemaate argued for "perpetual care" to make sure that law fulfills a responsible role in this expert knowledge domain.⁹²⁹ E.g. too much explicit consolidation of obligations and conceptualizations (of values, situations, and activities) is seen to pose a risk to the field's critical further development, but terms that leave too much interpretation up to the field can fail to force behavioral change, therewith undermining necessary development too—especially when obligations describe values that are more alive in larger society than they are in the medical field.⁹³⁰ Again, in a field with a strong history of resistance against the merits and practices of explanation, there seems to be reason for law not to be too shy.

The question is why the WGBO makes no mention of the broader conversations that it assumes to be, and as we saw indeed are, endorsed in professional standards. An example related to the Public Health vaccination policies for the novel COVID-19 vaccines played out in practice during the research on the chapter. It was not included in the chapter research but is added here, as an anecdotal argument towards more solid codification. Emerging information on side-effects for different groups led to several sudden policy changes whose choices were not very well elaborated (if at all.) In the eyes of many Dutch GP's such shifts were haphazard, ill-informed, and sometimes

928 Section 5.3.3.4

929 Section 5.3.4.2

930 Section 5.3.2.1

dangerous in how they kept patients from vaccinations that could save their lives.⁹³¹ This was meant both in the literal sense (sudden vaccine plan reversals) and through fostering distrust in the medical sciences and the government itself by explaining next to nothing and offering no other sources either. In one instance, many GPs refused to comply (i.e. they continued to vaccinate).⁹³² They publicly justified their stance in news media, arguing that they could and would make their own risk-assessments based on their own engagement with available information, and that they would discuss these with their patients in a way that allows for responsible shared decision making: on the basis of honest explanation and information and their ability to sustain their patients' reasoning. They reported that they had a hard time 'winning back' their patients' trust, which they commenced to do by talking to them about how the vaccines had been developed.⁹³³ They were also critical about how (other) public vaccination bodies had omitted engaging preemptively with groups that are known to have low trust in public vaccination schemes. In conversations with these groups, justification of medical knowledge needs to acknowledge historical wrongs in medical knowledge practices since this is a major factor in many groups' lack of trust.⁹³⁴ Engaging with patient trust this way is the terrain of element two, but it needs preparation, which is the terrain of element one. To return to the question of why the WGBO does not engage with many of element one's values, this might simply express the lawmakers' lack of awareness, or reluctance to act on the knowledge of, the politics of medical knowledge and how some groups in society receive worse care than others because of it. It would be in line with a lack of action with regard to other domains of discriminatory dimensions of Dutch society. This lack is currently being addressed on government level, and an investigation of the care domain is named explicitly.⁹³⁵

931 Michiel van der Geest, 'Huisartsen ontsteld over vaccinatie-advies: "Niet verwacht dat Nederland hier zo klungelig mee om zou gaan"', *de Volkskrant*, 12 April 2021, sec. Topverhalen vandaag, <https://www.volkskrant.nl/gs-bba4245c>; Dorothee Hafkenscheid, "'Mensen hebben het recht om te vragen om het AstraZeneca-vaccin'", *Medisch Contact* (blog), last consulted 9 December 2022, <https://www.medischcontact.nl/nieuws/laatste-nieuws/vandaag-op-de-werkvloer-1/werkvloer/mensen-hebben-het-recht-om-te-vragen-om-het-astrazeneca-vaccin.htm>.

932 Malika Sevil and Jop van Kempen, 'Amsterdamse huisartsen prikken soms door: "Anders is het mensonterend"', *Het Parool*, 14 April 2021, sec. Amsterdam, <https://www.parool.nl/gs-b2c8a55a>.

933 Cited newspaper articles and personal conversations with two General Practitioners.

934 Section 3.3.1.3, similar missed opportunities were reported from other European countries.

935 Ministerie van Binnenlandse Zaken, 'Besluit van 3 May 2022, houdende instelling van een staatscommissie tegen discriminatie en racisme' (Ministerie van Binnenlandse Zaken en Koninkrijksrelaties,); Parlementaire onderzoekscommissie effectiviteit antidiscriminiewetgeving, 'Gelijk recht doen: Een parlementair onderzoek naar de mogelijkheden van de wetgever om discriminatie tegen te gaan'.

5.4.3 Element 2: engaging with the social-epistemic positions of explainees

Explainers are obliged to investigate the social-epistemic positions of explainees in relation to the decision-making modalities and underlying (input) knowledge at hand; can explainees be expected to responsibly provide (or have provided) the necessary input, and understand the output? If no (or can't investigate), rebel.

This element, like element one, obliges to 'prepare the table' for the negotiation of the how's and why's of decisional outcomes. This time the focus is on how explainees *will be able to* experience a just testimonial process. Explainers need to be able to demonstrate engagement with their explainees social-epistemic situatedness (on individual and group levels) with regard to the larger decisional process and methods: 'the system.' This includes engagement with how a system historically treated explainees as a group and individually. The need to rebel exists when explainers feel their explainees are in no position to participate in the decisional process responsibly.

5.4.3.1 *The need to go beyond understanding 'on behalf of' patients: support for element two in ethical and professional discourse*

The chapter discussed how very different takes on element two's themes are to be found in medical-ethical traditions that inform the professional standards – important as Dutch law (again) leaves much of what element two promotes up to these other fields of governance. The main focus was on the difference between (strictly) individual, and more situated, or relational understandings of patient autonomy. Faden & Beauchamp dismissed most socioeconomic, choice-influencing circumstances as 'questions of justice' that 'lie outside the informed consent relationship.' Contrarily, O'Neill argued for a description of informed consent's purpose as 'the reasonable assurance that patients are neither deceived nor coerced, and can judge for themselves that they aren't.'⁹³⁶ Her take expresses support for element two, Faden and Beauchamp's does not. This is not surprising, as the Model too starts from a relational understanding of autonomy.

The need for doctors to engage with their patients' social-epistemic health (care) situatedness in order to serve them well was also reflected in the chapters' discussion of the 'sociality of diagnosis.' Patients' understanding of what bothers them, what brings them to seek a doctor's help, are inevitably socially, and societally, informed. In light of the social dimensions and consequences of diagnostic notions and labels, GP's need to engage with patients' self-understanding and situatedness in this context. One example that was named was the 'black lung' disease. Doctors in the US were instrumental in responding to commercially informed diagnostic methods of mining companies, with methods that more fairly diagnosed their patients' disabilities.⁹³⁷

⁹³⁶ Section 5.3.3.2

⁹³⁷ Section 5.2.1.2

Put differently, a GP needs understanding of their patients' self-understanding in the larger health care domain to responsibly diagnose, treat, or refer them to the right entry points in the larger maze of specialist diagnostics – and to serve their information positions in the process of doing this. GPs also have an important role in the further development of public health care. As critical actors in the larger sphere of Dutch health care and social care policies, they are relied on to fend for beneficial 'practicing ways.' (e.g., GPs have raised their voices against policy makers' unrealistic assumptions of a highly autonomous, 'self-sufficient' citizen-patient population.) Relatedly, they need to be aware of societal tendencies to 'medicalize' problems of citizens, and remain critical of the promotion of medical treatment when the origins and solutions to patients' complaints are possibly better looked for in, or at least should also be seen in relation to, societal spheres (the ADHD example.)⁹³⁸

But genuine engagement on these points means that explainers need the incentive and capabilities to learn *from* their explainees' experiences, and not just *about* them—or at least to learn about them from other points of view (e.g., through qualitative empirical research.) In other words, medical explainers need to understand medical (knowledge) practices *in terms of* how patients experience them and what they mean for their lives.⁹³⁹ It will be hard to assess patients' ability to responsibly participate in medical decision making otherwise. And, as was also discussed under element one, a very long-term culture of authoritarian practices means that a lot of bridges needed to be constructed for this to happen. This was no different for Dutch GPs who were historically, strongly engaged with their patients' societal circumstances.⁹⁴⁰ For decades (and centuries before it), patients' well-being was not considered to depend on their grasp of a health incident but on their obedience. Their medical knowledge and understanding was not catered to.⁹⁴¹ When informed consent had become ethically and professionally established, and it was accepted that patients needed information to act in their best interests, GPs' understanding of explanation as an interactive process lagged behind. Patients' accounts were not considered informed enough to be of much use to themselves or GPs, and GPs continued to 'inform themselves' *about* patients.⁹⁴² Furthermore, the 'medical is technical, and the technical can't be explained' argument was used to justify against accusations of (continued) paternalism. Notwithstanding the arguments that can be pitted against this (see the previous element), GPs had a point: patients indeed lacked knowledge. They had also been trained to accept (or to not go against) authority, and to understand themselves as medically ignorant.⁹⁴³ GPs did express the wish that their patients would become more critical 'health consumers,' but they were right where they worried that 'simply' start treating them as critical,

⁹³⁸ Sections 5.2.2.2, 5.2.2.3

⁹³⁹ Sections 5.2.3.2, 5.2.3.3

⁹⁴⁰ Section 5.2.2.1

⁹⁴¹ Section 5.2.3.3

⁹⁴² Section 5.2.2.1

⁹⁴³ Section 5.2.3.3

autonomous decision makers under such circumstances would be, well, un-careful.⁹⁴⁴ In a sense, this is ‘rebellious,’ but they are their own addressees. The next section looks at law’s ambitions with regard to improving this situation: to support doctors in realizing these aims, and to correct those who don’t even try.

5.4.3.2 Understanding the process from patients’ point of view: what guidance from law?

The WGBO provisions don’t explicitly engage with patients’ social-epistemic information positions, unless it expects all of these to express naturally in patients’ individual ‘situation’ and ‘needs.’ Doctors are obliged to inform themselves about these (how this is a problematic framing was discussed above) so that they in turn can inform patients about what they should ‘reasonably know’ about diagnostics and treatment options. The WGBO does not oblige that care professionals verify their patients’ understanding of this information. And in deciding what ‘informing’ amounts to in terms of content, the Handbook writes that the benchmark is an abstract: what a ‘reasonable person’ would consider in the context of (their) medically-informed decision making.⁹⁴⁵ Whether this refers to what they would *want* to consider, what they *need* to consider, or what they *should* consider is not clear. This leaves the bulk of choices with regard to the second element’s obligations up to doctors. The chapter also cited how disciplinary and public courts seemed to ‘side with doctors’ in how they tended to ignore research that shows that patients typically wish to be *more* informed than they are by their doctors, specifically about risks.⁹⁴⁶

The chapter’s findings include some relevant critiques from health law, and health ethics scholars (GPs among them) on the lack of context-sensitivity that expresses in the WGBO, and on what appears to be lawmakers’ lack of courage to use law to “realize the best possible health care.”⁹⁴⁷ E.g. ‘patients’ are made out to be capable, self-managing, autonomous individuals; contractual partners even in what in reality is a highly unequal relationship with their doctor in terms of information positions and social powers. Some were critical about how the WGBO’s contract-law character introduces assumptions of ‘horizontal relations’ whereas patients’ options are significantly influenced by other fields such as public policy, big pharma, insurers, decisions in and on medical research.⁹⁴⁸ To add here is how all these fields also influence patients’ medical self-understanding.

944 Section 5.2.3.3

945 Section 5.3.3.3

946 Section 5.3.3.3

947 Section 5.3.4.2

948 Section 5.3.2.1, 5.2.2.2

Legemaate was especially critical of the lawmakers' decision to *not* codify the obligation to verify patient understanding, especially since this is a solid part of the professional SDM standard that the law was meant to strengthen compliance with.⁹⁴⁹ The need for such legal support is amplified in light of how even the basic informed consent standards were ignored in practice until (and after) they were codified. And as checking patients' understanding is established to be hard to do, and to cost additional time in a practice that is under much time pressure already there may be even more reason to think this will be avoided. The medical handbook itself mentioned how it is hard to distinguish whether patients' complaints are about the contents of care, or the quality of informational relations with their doctors – a finding that does not attest to ideals of responsible and safe patient participation.⁹⁵⁰

News media's revelations of discriminatory biases in medical research, policy and practices (including unsafe medication) are of relevance for element two as well. It will be harder for patients to 'grow' in terms of social-epistemic information positions if they don't know about such issues, or don't feel safe to speak with their GPs about 'the politics of medicine.' When GPs don't learn of such concerns, this will leave them in the blind about e.g. their patients' trust and therapeutic compliance.

The space for law to consolidate more of element two's objectives than it currently does is arguably large: the GP field's (self-)definitions, descriptions and educational texts clearly support what element two promotes, and the KNMG's WGBO implementation guidance writes that the GP-patient relationship is considered exemplary for the larger field.⁹⁵¹ GP education materials for example emphasize the need for GPs to engage with (groups of) patients' social environments, social-epistemic positions, and information needs. As 'central head quarters' of patients' diverse health care relations and situations. GPs' professional norms oblige them to 'bring out their [patients' contextual] needs and values and verify their patients' responsible understanding.'⁹⁵² Patients cannot be relied on to mention an important value or factor if they don't know that it has medical relevance. Bringing such values out are qualitative skills that are explicitly named in GP education materials over various skill descriptions. But patients and their peers have no knowledge of such norms. Legal uptake gives them a stronger position. A public campaign *about* such a legal uptake could support the GP-patient relationship by 'breaking the ice,' and the possibility to check compliance could have a positive effect on any extant discomfort, unwillingness or unawareness on medical professionals' side.

949 Section 5.3.4.2

950 Section 5.3.4.3

951 Sections 5.2.2.2, 5.3.2.2

952 Section 5.2.2.3, 5.3.4.2

5.4.4 Element three: practicing interactional justice

Explainers are obliged to practice interactional justice, which entails to recognize explainees as knowers and rights-holders. Explainees should be provided information that is proportionate to their pre-investigated and incidental (self-expressed) needs; their knowledge and understanding of relevant, larger & smaller knowledge making processes at hand should be discussed with them with the aim of promoting their responsible (dis)trust; accessible justificatory sources from outside of the authoritative setting need to be pointed out accompanied by instructions on how to follow up on such leads; explainees need to be afforded information about their rights with regard to the explanation and the decision outcome; the possibility of social pressure needs to be mitigated by e.g. allowing to bring allies or make recordings.

The duties of this element describe the interactional dimension and behaviors that need to be given an explicit place in the testimonial process. If any description goes beyond what a process is seen to need, this will need to be justified in the testimonial record. The inclination of lawmakers to treat much practiced (or ‘bulk’) decisional processes as simple, self-evident, ‘routine’ and predictable has led to sub-optimal explanation practices. The implementation of automation in such cases exacerbates the problems while obscuring their origins.

*

5.4.4.1 Governance of social-epistemic inequality in SDM: a critical need

The Model’s understanding of explanation practices as knowledge making practices expresses strongly in element three. This makes the element of especially high relevance in a domain where shared knowledge making has become the point—even if, or perhaps, because, this is officially expressed as ‘shared decision making.’ The first section discusses recognition of this based on especially the non-legal studied literature, the second section again treats the WGBO’s expression and address.

Katz’s arguments, published a decade before the WGBO was enacted, deal with element three’s objectives explicitly: “in the absence of any one clear road to well-being, identity of interests [between doctor and patient] cannot be assumed, and consensus on goals, let alone on which paths to follow, can only be accomplished through conversation.”⁹⁵³ He advocated for explicit obligations for doctors to develop the necessary capabilities, to acknowledge the limitations of their knowledge, to learn to engage with, and to trust their patients’ reflective capabilities.⁹⁵⁴ O’Neill’s description of the purpose of informed consent as ‘the reasonable assurance that

⁹⁵³ Katz, *The Silent World of Doctor and Patient*. 1984, xlv, 102 section 5.2.3.3.

⁹⁵⁴ Section 5.2.3.3

patients are neither deceived nor coerced, and can judge for themselves that they aren't' is equally pertinent.⁹⁵⁵

For the highly unequal doctor-patient relationship, this translates into a need for GPs to de-emphasize their own authority in two ways: epistemically and socially. The need for this is sustained by the field's historical developments. The chapter described several historical shifts of focus in Dutch GPs' relations with patients: GPs zoomed in on, and out from, social and societal factors of influence on their patients' states; they prioritized, then deprioritized psychological states and causes, and eventually put patients in the spotlight at the start of consultations by letting them declare the reason of their visit. Along the way, the adverse effects of GPs' highly personal care bonds, in combination with assumptions of insight that in reality lacked patient-side wisdom had backfired; diagnoses were missed and patients social-epistemic needs were ignored to the point that expressing what they needed help with was argued to be irresponsible.⁹⁵⁶ In terms of element three, the lack of interaction with patients as 'knowers and rights holders' had resulted in poorer medical knowledge making about them, and in 'coercive practices.'

Contemporary GP (ethical and professional) standards paint a wholly different picture. Studied materials clearly emphasize patients' interactive informational needs. The focus is on the development and perpetual maintenance of a responsible, trusting, working relationship in which GPs remain in dialogue with patients. They need to be attentive to their patients' contexts and backgrounds, to bring out their social as well as informational needs and values. Informing *per se* is acknowledged to be therapeutically beneficial, and responsible patient understanding is described as instrumental to the perpetual development of a responsible trust relationship.⁹⁵⁷ The earlier element's suggestion to acknowledge the existence and avoidance of discriminatory practices is relevant here too.

The chapter also cited a type of 'moral consultation' for GPs to discuss particularly value-laden medical situations with their patients. The method entailed a broadening of the deliberative 'space' with other knowers such as (in)formal carers, and other 'knowledges' such as norms from e.g. law, ethics & professional standards, and health policy.⁹⁵⁸ In other words: a setting that situates the doctor-patient relationship in the larger social-epistemic world it exists in, and in which there are other authoritative grounds for the medical decision that needs to be made are introduced: 'out-roads from authority,' in terms of element three.

955 O'Neill, 'Accountability, Trust and Informed Consent in Medical Practice and Research' and section 5.3.3.2.

956 Sections 5.2.2.1, 5.2.2.2

957 Section 5.3.2.2, 5.3.4.2

958 Section 5.3.2.2

The inclusion of informal carers could however, and arguably should, use more promotion than was found in the studied literature. For the knowledge making that needs to happen between doctor and patient, patients also need a ‘safe space’ that lets them bring their questions, concerns, experiences, needs and wishes forward. Legemaate indeed suggested that guides or coaches could support patients in SDM practices, and a report from the new Independent Complaints Officer suggests room for improvement on this point as well.⁹⁵⁹ Many patients had reported that they were uncomfortable to speak with their GPs directly. The presence of ‘extra eyes and ears’ could help to prevent that the ‘reconciliation’ process fails to support the less powerful parties’ interests. Moreover, *expert* guides could help to tease out any social-epistemic entanglements that underlay the complaint, itself. As the Handbook also pointed out, complaints about the ‘content of care’ and ‘care communication’ were hard to separate. Another clue exists in the KNMG’s recommendation for doctors to ask their patients to tape conversations, or bring a trusted person to consultations. Neither has become common practice, and taping conversations was reportedly unpopular.⁹⁶⁰

5.4.4.2 *Social-epistemic interaction in the WGBO: little support or.. potentially undermining?*

The first (1994) WGBO codification strengthened patient’s informational positions by codifying their right to a limited set of informational categories, but simultaneously strengthened doctors’ epistemic authority by leaving most interpretation and explication up to them and framing the relationship as one of theoretical equals.⁹⁶¹ In light of the medical field’s history and those of the modeled duties of explanation care, this ‘setup’ arguably expresses a very weak promotion of the type of relationship that was called for. Legal evaluations indeed showed how the law did not manage to force much progress on GPs’ information paternalism. Around the millennium a decline of patients talking time in GP consultations was reported. Some more and less ‘acceptable’ GP justifications were cited before: the technological complexity argument, EBM influences, the lack of patients’ knowledgeability, the non-representativeness of the provisions (‘isolating’ explanation would disrupt the natural process). Legemaate’s response to the latter was cited: without at least pointing out what should at least be achieved at a certain point in time, doctors’ control over patient information rights would simply remain total.⁹⁶²

With regard to the EBM argument, the thesis’s angle would add another consideration: if GPs could not be relied on to secure the place of their (qualitative) patient-centered methods themselves, is that something that fits the lawmaker’s cost-reductive aims, consecutive governments’ mainlining of individual patient resilience and autonomy, or

959 Sections 5.2.3.3, 5.3.4.3

960 Section 5.3.4.2

961 Section 5.3.3.2

962 Legemaate, *Goed recht*, 59 section 5.3.2.2

possibly both? Cited critique on Dutch health law (more broadly) expresses tentative support for such suspicions. E.g., the paradigm is overly focused on informing independently reasoning patients' choices *for* treatment, rather than on supporting persons in vulnerable states' social-epistemic needs for a type of decision making that is inherently ambivalent. Guidance for other kinds of 'choices' such as abstinence from treatment are lacking, informational needs with regard to self-care, self-monitoring, and relations with informal carers are unnamed. Moreover, the law seems to assume that 'bits of information' that patients' choices turn on are identifiable: patients only have a claim about being under-informed if they can prove the information would have changed their decision.⁹⁶³

These criticisms were arguably not ameliorated when the WGBO's information obligations were finally updated in 2020 to consolidate (aims of) SDM practice. As was discussed, some pushback of paternalism is visible in the changes. All patients (not just the young and less able) now need to be informed in ways that 'are appropriate to their comprehension,' some more informational content is added, authoritarianism curbed, as reflected in terms like 'proposed treatment' instead of 'treatment that the care provider regards as necessary.' Added are obligations to discuss options and alternative options, to invite patients' questions, to engage with their situation and needs, and (upon request) provide written information.⁹⁶⁴ Still the new rules leave a lot unsaid and unrequired, such as the earlier named omission of checking patient understanding. For element three, and in light of the complexity of medical knowledge and the historical information imbalance between the parties, the word 'explanation' is saliently lacking. And in light of the field's historical instances of coercion, authoritarian traditions, and resistance to information duties, explaining what the information duties are for could have been an option: it would respect patients as rights-holders. Think of the 'black scenario' playbook for IC ward scarcity in the COVID-19 pandemic. Members of IC wards decision-making teams were uncomfortable to face patients and asked to remain anonymous – which would (illegally) leave a life-or-death decision unaccounted for.⁹⁶⁵ The chapter referred to critique that patients have a right to know who makes decisions about them, but to add here is that they also have a right to know *why*, and that they need to know this. And medical decision makers need to know that they know this: a patient is arguably better served by a decision maker who refuses to act, rather than 'go stealth,' if they fundamentally disagree with the justness of their act.

The question becomes whether such minimal explanation governance could possibly *undermine* a beneficial practice, and not just fail to support its progress. E.g., a law that repeatedly fails to express certain societal needs arguably expresses and supports the opposite. Concerns with regard to hard-to-attain ideals of SDM practice, which risk to be obfuscated by a simplistic legal representation, spring to mind. Named were hard

963 Section 5.3.3.3

964 Section 5.3.4.2

965 Section 5.3.2.1

to bridge knowledge and understanding gaps; inevitable social power asymmetries that make patients into bad self-representatives; unclarity about the information patients need to have brought to the table to establish that the decision was in fact ‘shared.’⁹⁶⁶

5.4.5 Element 4: creating records

Explainers need to create records of explanation practices. These should be understood as truthful accounts of the testimonial exchange as it was prescribed under element three. Therewith the record should express how all previously described duties were attended to, or provide reasons for when they were not. The records need to be shared with explainees, and made available for outside scrutiny in accordance with rules that govern the decisional domain and relevant privacy and data protection regulation.

These record-related duties are meant to produce more comprehensive accounts than the ‘statements of reasons’ that are typically the outcome of decisional processes. This acknowledges how explanation is a knowledge making practice itself, and therewith a place or conduit of possible oppression. Comprehensive records can sustain progressive development of decision and explanation practices across time and domains.

*

At face value, this element seems ‘under served’ in this domain. Per the rules, patients are for example only given written (including electronic) information upon their request, and ‘records of explanations’ are not mentioned at all. The consequence is that patients don’t automatically exit consultations with something they can share within their social circles and with informal carers, whereas there is an acknowledged need for patients to receive help with processing diagnostic information (as well as instructive information, see the ignored KNMG recommendation above.) Support for more record creation can also be found in Legemaate’s suggestion that a legal obligation to verify patient understanding would allow to assess doctors’ compliance with this important professional-ethical norm. Having records could support such investigations, in addition to more labor-intensive empirical studies.

Some records of what patients are told will be in patients’ medical files, since these also serve as evidence of the obligatory informed consent process. Time constraints did not allow to trace medical file related obligations. There is a small revolution going on with regard to these files. Care providers are starting to move patient files online in a way that patients can access them remotely. Not discussed in the chapter was a new problem that arises with the implementation of ‘real-time’ remote patient access to for example diagnostic results. Various doctors voiced concerns about emotional and ‘cognitive’ risks that arise when patients are confronted with impactful findings that are hard to understand without their (or ‘a’) doctors’ qualification. They argued for

⁹⁶⁶ Section 5.2.3.3

a ‘pause button’ under doctors’ control, but in reply, the director of the Dutch Patient Federation argued the use of such a button should be a patients’ choice.⁹⁶⁷

Another use for record creation that exists ‘once removed’ from explanation practices can be identified in the context of informal complaints procedures for GP practices. In addition to individual needs (e.g., ‘to feel heard’) patients reported to hope that others would benefit from any improvement that would result from the resolution of their complaint. To this end, patients were eager to be informed about the uptake of their complaint, and disappointed as such feedback was rarely shared. The chapter discussed how the aim of the procedures was ‘reconciliation’ rather than resolution, but since complaints about the *content* of care and the *communication* of care were hard to separate, a ‘reconciliation session’ might well validate as an ‘explanation session’ and be usefully brought under explanation rules.⁹⁶⁸

5.5 Chapter 5 in a nutshell

This chapter evaluated the main legal explanation rules for Dutch General Practitioners. It placed them in relation to the field’s self-governance, taking on board a selection of explanation related norms from professional and ethical domains.

Much explicit recognition for the modeled aims and values was found in studied literature on subjects like the nature of medical knowledge, the domain’s relatively recent move away from purely authoritarian decision making, and the daunting task of responsibly sharing decisions with patients: the contemporary norm. The Model-based analysis allowed to form an opinion on the ambitions of the Dutch lawmaker with regard to the guidance of doctors and patients towards this paradigm. More abstractly speaking, it allowed to form an opinion about the explanation related guidance of a highly unequal social-epistemic relationship in an expertise based (as opposed to a rule based) domain that caters to vulnerable subjects.

These ambitions were argued to be modest, or rather, limited. One consequence of this is that in light of the many Model-pertinent subjects that the legal rules leave unaddressed, the choices that law does make can become unhelpful beacons (such as when it deprioritizes ‘non-instrumental’ informational exchanges), or invite doctors to ignore the rules altogether (on the ground that these are ‘disconnected’ from actual explanation practice.) Below, a brief summary of important points of recognition is followed by a brief characterization of law’s engagement. Like in the preceding domain’s chapter, this ‘nutshell’ does not summarize all the chapters findings but rather summarizes the usefulness of the analysis. The chapter after this one relates the

⁹⁶⁷ Stephanie A Kooiman, ‘Realtime-inzage via het patiëntenportaal’, *Nederlands Tijdschrift voor Geneeskunde*, 4.

⁹⁶⁸ Section 5.3.4.3

analysis to AI-infused times to argue the need for explicit legal explanation guidance over softer approaches.

In general, the historical absence of explanation obligations was reported to have had profound adverse consequences for the quality of medical knowledge development and the fair and safe treatment of patients. Fast forward to the 1960's in which this situation was under increasing criticism, it was discussed how this authoritarian, paternalistic 'spell' was not easily broken. This was no different for the very intimate and traditionally long-term bonds of Dutch GPs and their patients. There was a persistent lack of interaction with patients as 'knowers and rights holders.'

Both GPs and patients were ill-prepared for the more honest, mutually well-informed relations that were argued for from within and outwith the field. A persistent 'excuse' (used internationally) for not obtaining such relations was and is grounded on arguments that understanding medical knowledge will always be unfathomable for patients. The excuse was argued to obfuscate deficiencies in doctors' own capabilities (e.g., to learn from their patients, rather than 'obtain' knowledge about them, and to understand what information is meaningful for patients), to ignore the politics of so-called 'technical' medical knowledge, and the plight to settle the score of centuries of patient 'ignorance making.'

A wholly different approach was found in contemporary medical ethical-professional norms for the GP relationship. GPs as well as their patients are described in terms of their social-epistemic positions; medical knowledge is described as decidedly social and as requiring qualitative approaches alongside other types of knowledge making. GPs are expected to be critical, ism-aware scholars, contributors and practitioners. They need to understand their patients in their social-economic-cultural context, and verify their patient's responsible medical understanding. Informing patients well and continuously is described as a fundamental precondition for a responsible trust relationship in which decisions can be shared responsibly.

Still some Model-pertinent aspects are non-explicit in these descriptions, such as the need to discuss historical and contemporary wrongs of medicine, or the sociality of medical knowledge in general. Such conversations are important for groups of patients whose trust is either understandably low or the opposite, too high. In light of the field's track record, it would be naive to expect that such conversations establish naturally or that all doctors are able to conduct them. In addition, very different understandings of what non-oppressive patient consent turns on, or depends on, exist in international medical ethical discourse. Individual autonomy versus relational approaches make for very different outcomes in terms of explanation practices.

It also became clear that proper professional norms don't necessarily translate into any kind of explanation practices—compliance with informed consent obligations was found to be lacking for decades in the GP domain, even after (partial) codification, and even after this law had gone through an elaborate implementation process to improve compliance.

These observations raise questions and expectations with regard to law's ever-developing guidance of the explanation relationship between what will always be very unequal parties, both socially and epistemically speaking. It is hard to say how it is best established that a decision between them was indeed 'shared', but it will be harder when the norms for sharing decisions responsibly are mostly made, and only known, by the more powerful party. Legal rules, at least, are made through democratic processes with which all parties theoretically have a voice, and they are made public to all whom they pertain to. The fact that this itself is a very criticizable idealization makes it all the more important to see what 'medical explanation rules' at least try to do.

Several points of criticism were discussed in the chapter, starting with how the law casts the doctor-patient relationship as one of contract, ignoring their highly unequal social and epistemic positions and patients' inherent vulnerability. Within this frame patients are cast as capable, self-managing, autonomous individuals, in line with general tendencies in contemporary Dutch care and welfare laws.

The law, for example, speaks of patients' needs, not values, concerns, or doubts. Doctors are obliged to 'inform themselves' about these, ignoring how 'needs' are typically articulated along the path of mutual informational exchanges and influenced by what doctors decide to share, and ask. The absence of the term 'explanation' is interesting to note in this respect, as is the need to check patients' understanding.

More generally the legal paradigm assumes the value of any particular information to depend on how it explicitly factors in patients' decision making. A choice that claims to serve patients' 'freedom to choose' while ignoring, firstly, established insights about how patients' decisions are inherently ambivalent and don't necessarily turn on any piece of information. The vision also ignores the value of explanation per se, also for treatment that will be consented to no matter what, or in cases of choices for non-treatment. In legal commentaries, the governance of discussions about profound medical 'dilemmas' like those posed by novel diagnostic affordances are entirely shied away from as belonging to the domain of professional standard setting. That choice is questionable in a field whose practice spheres and cultures are historically explanation-averse and where racist and discriminatory dynamics are acknowledged to be very large—a point that cannot be ignored in light of how these novel techniques are increasingly AI powered and therewith invite inscrutability as well as exacerbation of wrongful knowledge making. GPs, arguably, could use the backup of strong explanation obligations to demand insightful and fair medical knowledge to 'deal with.'

Finally, the fact that there is no legal, or (as far as was researched) other standard obligation to give GPs' patients an explanation record was considered. There are certainly arguments to be made for such records, and for research access to (aggregated) records but this particular point needs further investigation. The governance of medical files needs to be investigated to that end, and the same is true for the relatively novel practice of electronic patient files.

Care to explain?

6 Toward care-ful legal explanation regulation: lessons from the present, for the present

6.1 Looking back, looking forward

6.1.1 Drawing lessons from the domain studies

In response to fundamental, explanation-related challenges of ADS and AI, various legal efforts to safeguard established aims of explanation are underway. Relevant points of attention are alerted to, and tended to, in these efforts. The same is true for the multidisciplinary research that sustains, promotes, and criticizes such efforts. This thesis engaged with prominent concerns about how decision practices are opaquely influenced, and influenced by opaquely wrongful, knowledge practices; it engaged with the non-individual nature of harms that ensue, and how the bafflement that hits explainees as well as their designated explainers poses obstacles to responsible participation in decision making. The thesis has argued that these are all valid points, yet they are also distracted. Through how the gaze of such efforts is strongly focused on intricacies of new knowledge making methods, the arguments reflect a false sheen of novelty with regard to the perceived explanation challenges. The consequence of this is that additional, AI-informed explanation rules are crafted as above-ground reparations on the premise that the fundamentals of explanation regulation are solid. Established legal explanation ideals continue to be appealed to as shaping norms for the new explanation practices, and to inform the review of practices that are seen to ‘succumb a bit’ under the pressure of ADM.

Acting on the basis of a grounded suspicion that our contemporary explanation ideals stand in false light, and that they *don't* pull their weight when it comes to protecting explainees from—especially—latent knowledge-related oppression, the previous two chapters investigated two basic, fundamental sets of explanation rules. Each treated a domain of different character: a rule-based decision domain, and an expertise-based decision domain. The fundamental explanation values of both domains are prominent in the above-described idealizations.

The tool it used for this investigation was the modeled duties of explanation care. The Model was built on insights from philosophical fields concerned with, bluntly summarized, rights & wrongs of knowledge practices. The chapter that describes the Model's construction approached explanation as ‘knowledge making about knowledge making,’ and explanation rules as behavioral instructions for explainers: as obligations, regardless of whether the actual rules of a domain are cast in terms of explainer duties

or explainee rights. For the thesis's purposes, this does not matter. It focuses on the human explainer as a person in a designated position of social and informational power who has the moral duty to, at the least, try and prevent to become an instrument of knowledge-related oppression—and more positively, has a role to play in furthering the justice of their practice, and the promotion of responsible knowledge spheres. At the time of writing, this person is pushed forward as the embodiment of humaneness in an uncaring machine age. At the same time, the sustainability of established explainers' roles are (rightly) questioned in terms of meaningfulness in AI-infused times. What can we learn from law's instructions?

To come to the purported investigation, the preceding domain research chapters presented a functional description of the decision making of each domain ('what, who, & how') followed by a discussion of the domains' main explanation rules in place. It then related the domains' decision making and explanation from the angle of the modeled duties of explanation care, tracing the Model's aims and values in the findings of the domains. This way, the ambition and potential of the explanation rules were assessed. 'Islands of recognition' for the Model's values and objectives were identified in the studied literature of each domain. Both with regard to the decision making and decision makers, as with regard to what needs to be explained about it by the latter. For example, both fields recognized the need to govern 'explanation' as a process rather than a discreet moment; the need to investigate the social-epistemic positions of explainees and make them matter; and a moral responsibility for explainers to engage deeply with the purported wisdom of their domain. But several observations revealed how both domains' explanation paradigms also frustrated these aims. Sometimes by shying away from making necessary obligations explicit, sometimes by ignoring established explanation challenges, sometimes because the lawmaker prioritizes ideals that are not as focused on 'wizening up.' In all cases, there was no shortage of knowledge available to the lawmaker that could promote a different course.

The effort therewith teased out several subjects and tensions that deserve attention in each domain. As lessons learned, these observations also deserve to be included in efforts that are being made towards 'meaningful' explanation practices across domains in AI-infused times. This is the objective of this last chapter. By discussing these lessons, the chapter explicates how the Model can support the work of explainers, researchers, and rule makers in times when shying away from discussing knowledge in explanation is (rightly) considered to be untenable. The chapter therewith answers the fourth research question: "which lessons from the analysis of existing explanation regulation should inform how we deal with ADM explanation regulation?"

6.1.2 Approach and structure

The chapter is structured in four sections that are each dedicated to one of the four modeled elements. Each discussion starts with a general reflection on the basis of observations whose recognition in both domains makes them of general weight.

After this, takeaways in the form of observations are discussed for each domain consecutively. The observations start from illustrations from the Model-based domain analyses. The illustrations are brief; the point is not to provide summaries of the previous chapters. Neither is the list of observations complete in the sense that all possible lessons are drawn in this chapter. For non-ADM regulation questions, the domain chapters themselves remain the primary source. The point of this chapter is to demonstrate how such analyses can (and should) inform contemporary efforts – a full-on ‘scan,’ complemented by empirical research is still recommended in the design of explanation governance for any particular domain.

The illustrations are related to ‘AI-infused times,’ with which the reflections are implicitly translated into recommendations. These parts are of interest to all readers, but the relevance of different observations will be of different weight for different actors (explainers, researchers, rule-makers, explanation designers). Some recommendations come in the form of things to take heed of, some are more directly instructive. In each element’s discussion, one or two illustrations from contemporary explanation literature are engaged with to sustain the translation from the domain analyses since these (generally) abstracted from this literature. The very last section of the thesis (section 6.6) takes this effort one step further, and grounds some more forward oriented thoughts on accumulated, technology-related insights of the preceding sections.

6.1.2.1 *A note on literature*

The chapter re-engages with AI-informed explanation governance, and as stated cites some additional literature. Among these are articles and reports about various forms of impact or risk assessment for ADM. Simply put, such assessments are tools of legal accountability regimes. Decision makers, system designers and/or providers perform such assessments to provide evidence that they have taken specified normative (legal, ethical) demands into account in their procurement or technological development. These norms may belong to the domain in which a system will be deployed, or belong to a certain domain of norms: think of Algorithmic, AI, Fundamental/Human Rights, Data Protection/Privacy Impact and Risk Assessments (AIAs, FRIAs, (D) PIAs, AIRA’s and variations). Impact and risk assessments are increasingly made obligatory or come strongly recommended in different legal frameworks as a

pre-employment and/or periodic, continuous condition for using ADS,⁹⁶⁹ and so approaches and methodologies for them are being developed widely.⁹⁷⁰

This thesis is not about decisional accountability in the straightforward legal sense, but it does describe *conditions* for accountability. It deals with account-ability,⁹⁷¹ or put differently, with accountability for the knowledge-related dimensions of a decisional practice. It is from that understanding that the chapter engages with this literature—or, arguably, that the literature is engaging with the thesis’s subject. Writers will be cited who alert to how information that can be created through assessment processes are a ‘meaningful’ resource for individual explanation practices and obligations. Whether this is so would also depend on other factors than the information itself. Not much is won when the assessment processes are opaque themselves; when they abstract from realistic context,⁹⁷² exclude affected parties or (other) important societal representatives,⁹⁷³ when they exhaust such parties’ resources to bring risks to the fore,⁹⁷⁴ misrepresent their interests,⁹⁷⁵ or when meaningful aspects remain unaddressed for other reasons. Not much is won at least for explanation practices directly. Documenting the processes will still be useful, as records to study about the kind of practices that are being developed, the societal values that express in them (or not) and who is responsible for that happening.⁹⁷⁶ It is in

969 See e.g. ‘Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data.’, Pub. L. No. 32016R0679, 119 OJ L (2016); ‘Regulation (EU) 2017/ 745 on Medical Devices’ (2017); ‘Proposal for a Regulation laying down harmonized rules on artificial intelligence’, Pub. L. No. COM(2021) 206 final (2021); It needs to be noted that as yet, the obligations in the EU draft AI act are criticized for their narrow scope ‘AlgorithmWatch’s Response to the European Commission’s Proposed Regulation on Artificial Intelligence – A Major Step with Major Gaps’, *AlgorithmWatch* (blog), April 2021, <https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021>.

970 For a ‘snapshot’ of approaches in 2021, see for example ‘A survey of artificial intelligence risk assessment methodologies: The global state of play and leading practices identified’ (EY and Trilateral Research, 2021).

971 Daniel Neyland, ‘Bearing Account-Able Witness to the Ethical Algorithmic System’ (2016) 41 *Science, Technology, & Human Values* 50.

972 Jacob Metcalf et al, ‘Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts’, in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT ’21 (New York, NY, USA: Association for Computing Machinery, 2021), 735–46, <https://doi.org/10.1145/3442188.3445935>.

973 Emanuel Moss et al, ‘Assembling Accountability: Algorithmic Impact Assessment for the Public Interest’ (Data & Society, 29 June 2021), 50.

974 Dillon Reisman et al, ‘Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability’ (AI Now Institute, April 2018), 20.

975 Mona Sloane et al ‘Participation is not a Design Fix for Machine Learning’ (arXiv, 11 August 2020), <https://doi.org/10.48550/arXiv.2007.02423>.

976 Selbst and Barocas, ‘The Intuitive Appeal of Explainable Machines’; See also Amore: analog to how Feinman asked all scientists who contributed to the launch of the Challenger to write down where their risk calculus was based on, how their piece of tech, or their knowledge of it might fail, “can we imagine today an equivalent method of asking the scientists we research to please ‘write the probability of the failure of your piece of the machine learning software on this piece of paper.’ How would the reinstatement of doubt account for the adjustments of weights that are conducted by the algorithm on itself?” Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, 147.

these discussions about the above-named caveats with regard to impact procedures, as well as in suggestions of meaningful information that the thesis identifies arguments that indeed testify to the usefulness of relating the two fields.

6.2 Investigating explainer authority (element one)

6.2.1 General observation

To reiterate, the first element obliges explainers to investigate their own social-epistemic positions to ensure that their understanding justifies their authoritative and trustworthy explainer positions. The point for them is to avoid to become instruments of unjust practices, and to be able to explain their efforts towards this aim: “the point at this stage is to link the self-reflection of explainers to their position of authority *vis-à-vis* explainees, as part of responsible practice.”

Examples of the need for decision makers to be critical practitioners, to be on the alert for innate or latent injustices of their decisional knowledges, methods, rules, and norms were identified in scholarly literature across the two domains. Advocacy for relating the investigative, reflective activities that sustain this aim to decision makers’ *explainer* roles was however less strong—although stronger in the GP context than in that of the Administration. In legal explanation rules of both domains, the expression of element one’s objectives was entirely lacking.

A notable difference in the studied explanation governance materials of the two domains was the extent to which it was acknowledged explicitly that knowledge-related oppression is an issue to be on the lookout for in the first place, regardless of whether this should be anchored in legal rules. E.g., the existence of biased knowledge practices was at least acknowledged in self-governing layers of the GP domain, but both administrative bodies and the responsible Ministries are struggling to admit the same happens in law-and policy making. This raises questions about the understanding, awareness, and intentions of the Dutch Lawmaker with regard to both domains. Below, domain-specific illustrations are given that should inform us moving forward.

6.2.2 Observations from the Administrative domain

6.2.2.1 *Resistance to progressive understanding*

Calls for an explicitly justice-oriented bureaucratic body, whose members engage in ‘meaningful relations’ with citizens, who will guard due processes and provide solid and understandable justifications are acutely amplified in the Benefits Scandal’s ‘crisis of constitutional democracy.’ But such ideals, and grounded notions about what needs to change institutionally to meet them had been advocated for much longer. In legal and

social research, ombudsman reports, at times in parliament, and more recently in the UN's address of the Dutch judiciary in relation to ADM. A telling anecdote: after the Scandal, the Dutch Prime Minister—responsible during the decades that the scandal established and unfolded and for the failure of restorative efforts so far at the time of writing—announced his 'radical idea' for a more oppression-resistant governance structure. He proposed to institute an independent, Government-funded investigative body that reports to government about problematic administrative practices (if his words to gather a "club" of people are euphemistically interpreted.) The National Ombudsman replied incredulously: such a club was instated in 1982, so perhaps a club that actually reads the Ombudsman's reports could be established?⁹⁷⁷

The lack of progress made by administrative bodies, and the reluctance to engage with established causes by those who are end-responsible should serve as warning for the National 'Algorithmic Watchdog' that is being established at the time of writing,⁹⁷⁸ for legislative employees working on ADM and AI accountability and explanation governance, for the newly instated commission on latent racism and discrimination in Dutch society,⁹⁷⁹ and the Senate initiative to strengthen racism and discrimination scrutiny for new and existing Dutch law to mitigate the limited reach of anti-discrimination law and legal safeguards.⁹⁸⁰ To force overdue 'progressive maintenance' and make their own future findings count, such bodies would do well to investigate why relevant insights, including research directly and indirectly commissioned by the State itself are not being made to bear. The domain research would advise to engage with the type of moral thinking, the type of 'value rationality' that administrative bodies are asked to do (and are assumedly trained for), also in relation to the norm setting of other *trias* branches. E.g., authors have voiced concerns that the lack of justice-serving push-back of the Administrative Judiciary expresses a lack of 'critical apparatus'.⁹⁸¹ The chapter however also discussed that when critical Judicial corrections of administrative decisions *are* issued, administrative bodies, i.e. the State, tend to file for appeal or even ignore the verdict. In other words, they tend to fend for their own policy interpretations—interpretations that are not under democratic scrutiny and that

977 Guus Valk, 'Nationale Ombudsman: "Laat Rutte maar een club oprichten die onze rapporten leest"', *NRC*, last consulted 19 November 2022, <https://www.nrc.nl/nieuws/2021/05/11/nationale-ombudsman-de-afrekencultuur-bestaat-nog-altijd-a4043283>.

978 'Coalitieakkoord 2021 – 2025 Omzien naar elkaar, vooruitkijken naar de toekomst' (VVD, D66, CDA en ChristenUnie, 15 December 2021).

979 Ministerie van Binnenlandse Zaken, 'Besluit van 3 May 2022, houdende instelling van een staatscommissie tegen discriminatie en racisme'.

980 Parlementaire onderzoekscommissie effectiviteit antidiscriminatiewetgeving, 'Gelijk recht doen: Een parlementair onderzoek naar de mogelijkheden van de wetgever om discriminatie tegen te gaan'.

981 Schuurmans, 'Toeslagenaffaire: outlier of symptoom van het systeem?', Causes named are Administrative legal traditions (that are being 'upheaved' at the time of writing) in combination with a prohibition for them to test primary laws for constitutionality.; Margreet Fogteloo, 'Hoogleraar mensenrechten Barbara Oomen: We zijn nonchalant over onze rechtsstaat', *De Groene Amsterdammer*, 30 June 2021.

were repeatedly shown to exacerbate, rather than ameliorate, a lack of justness extant in underlying laws.

6.2.2.2 *Unreasoned hardship*

Related to the above are observations about Administrative law's understanding of what amounts to 'wrongs' in the first place. A fundamental obligation in service of the prevention of injustices is the codified principle of proportionality: "[t]he negative consequences of a decision shall not be disproportionate relative to the objectives that are pursued by the decision." Discussions and confusions about why the clause was ignored by administrative bodies (and judiciary) in the Benefits Scandal are not repeated here, since recent jurisprudence developments have confirmed the 'always on' status of the principle, and in addition extended the assessment of compliance that the Administrative courts need to perform.

The question is, to what extent the stronger engagement with the principle will further administrative bodies' engagement with knowledge-related harms. The premise of the clause is that hardship originates on executive levels, rather than follows from legal objectives themselves. Scheltema, concerned that Benefits-like injustices can slip through the cracks, argued for further codification of the principle to clarify that rules themselves may be diverted from "to the extent that is needed to apply it in a more balanced fashion."⁹⁸² The suggestion was met by concerns of arbitrariness (see below under the next lemma for how this is a notable observation in itself) and superfluousness.

Where the domain analysis suggests that it is already naive to expect administrative bodies to ameliorate or tune down legal expressions of 'harsh political climates' in their policies and decision making, the framing of hardship as executive accidents is especially problematic in ADM times. Injustices creep in on all 'rule making' levels, and are becoming harder to identify. This situation can be exacerbated rather than ameliorated as a result of additional automated systems designed to select possible instances of hardship for individual, human review. Such methods of selection come with their own challenges, and may introduce additional injustices such as delays in eligibility decisions and effectuation,⁹⁸³ exacerbated loss of privacy that typically come with affordance scrutiny,⁹⁸⁴ and, as would for example apply in the phantom vehicle cases, the narrow definition of 'hardship' means that those just above the 'mark of compassion' still need to pay their large fines. The point of the hardship clause is that it is impossible to foresee all the ways that the application of a rule or law may lead to hardship, and using automation to assess that risk, like the assessment of other risks (as a typical output that AI systems are designed for) is seen to encourage human decision

982 Section 4.2.1.4

983 Binns, 'Human Judgment in Algorithmic Loops'.

984 Bridges, *The Poverty of Privacy Rights*.

makers to add weight to the factors that these systems (can) treat, to ‘think in terms of the system’ to the detriment of other possible considerations available to them.⁹⁸⁵

But there’s an additional dynamic to take note of. In the aftermath of the Benefits Scandal, some unjust treatment remained stubbornly unreasoned. Even when wrongful and disparate effects of a legal climate of distrust towards citizens in need of support were acknowledged, when unrealistic assumptions of self-sufficiency were acknowledged (again with disparate effects), the *surface-ability* of disparate and marginalizing effects were not. Neither were directly racist and other discriminatory policies of the Tax Administration. At the time of writing, the new State Secretary has finally acceded that there institutional racism was at play—immediately followed by the remark that since institutional racism is ‘not a ‘legal concept,’ no damage claims were expected by the Ministry: only proven instances of individual racism and discrimination would be awarded.⁹⁸⁶ The combination of such authoritative instructions with the persistent denial of institutional and clearly racist wrongs make it hard to see what ‘deep’ engagement with unjust potential of laws, rules, and decisional systems is to be expected without clear codification about the type of oppressive knowledges that need to be accounted for.

6.2.2.3 *Due diligence and motivation: a case for (more) interdependence*

As—especially—the Vehicle License Registration cases showed, citizens who were left ‘out in the cold’ were generally also left ‘out of the know’ before, during, and after decisional practices. This exacerbated their situations since they were ‘set up to fail’ to participate in their own interest. ADM-inspired calls to improve (or prevent) such practices tend to name the further development of the principles of motivation and that of due diligence together. Their combination certainly fits with Element one’s aims, since it links decision makers’ investigative burdens to their obligations as explainers. The domain analysis however also called attention to how in the Awb’s explanation paradigm, the two sides are not that strongly related in reality. This needs attention in light of the promoted *further* development of the principles.

985 Ben Green and Yiling Chen, ‘Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts’, *Proceedings of the ACM on Human-Computer Interaction* 5, nr. CSCW2 (18 October 2021): 418:1-418:33, <https://doi.org/10.1145/3479562>; See also Amoores, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, 55.

986 As argued by Bot, the letter testifies to a persistent reluctance in Dutch politics (and laws) to adopt internationally accepted definitions of racism and institutional racism as layed down in treaty law signed up to by The Netherlands. Citing Gloria Wekker’s ‘White Innocence: Paradoxes of Colonialism and Race’ (Duke University Press 2016), Bot qualifies the Secretary’s letter as a perfect example of it, a “militant renouncement of any occurrence of racism in The Netherlands” Bot, ‘Is institutioneel racismisme echt racistisch?’

The problem starts with legal restrictions on what needs to be justified. Many preliminary investigative activities are exempt, limiting what decision makers are obliged to think about in terms of justification. In other words, here we have a law that establishes precisely that which ADM-focused explanation rules ascribe to the implementation of technology. To be sure, it is certainly problematic that the specialty of AI is to produce obscurely unfair ‘predictive’ clues that lead to human investigations, and that automation of the larger decisional process reduces possibilities for due process checks and balances. But where it is already hard to make automation focused protections ‘work’ in contexts where the function and consequences of human (or ‘manual’) and automated steps need to be understood in their interrelation (and playing around with where to locate either kind of step supports legal ‘escape artists’),⁹⁸⁷ such solutions’ reliance on the fundament, the humaneness of ‘analog’ legal explanation paradigms in place, is arguably as ill-conceived. Of interest at the time of writing is that the Court of Amsterdam decided how a print-screen of an automatically generated decision in an online application form, stating how the applicant did not qualify for what they applied for on the basis of their income *and therefore the form would be aborted before the applicant could submit the application*, was to be considered an appealable decision. A small revolution in the making? It needs to be seen whether the State appeals.⁹⁸⁸

To continue, various Awb ‘system intricacies’ make various of the investigative engagements that are called for either hard to do, only ‘halfway legal,’ or the results hard to justify. With regard to practical-legal obstacles that make it ‘hard to do,’ think for example of the ‘iron cage’ of unhelpful information architecture, and the vast Administrative legal landscapes’ detailed complexity. An example of ‘halfway legal’: there is a codified principled prohibition to engage with other interests than those of the law that the decisional authority derives from, but the due diligence principle holds that ‘all information and interests must be allowed to count.’ Administrative bodies tend to create policy rules to exempt them from having to reason choices in this respect (with varying success). Lastly, things like the ‘single authority’ information paradigm, which obliges Administrative bodies to use designated (other) bodies’ fact establishment about decision subjects even when they know these to be incorrect lead to decisions that are certainly justifiable—but only on a ‘blame the system’ basis. Again, such observations suggest that without intervention, the Awb’s ‘due diligence and motivation’ paradigm aligns with problematic ADM practices whereas it is called out as an inspiration of the opposite in AI governance statements.

987 Binns and Veale, ‘Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR’.

988 ECLI:NL:RBAMS:2022:3066 Geautomatiseerde afwijzing na invullen online vragenlijst is bestuursrechtelijk besluit.

Lastly, when explainers provide bad reasons, this is not seen as an invalidation of the reasoned decision per se. This is a subject in itself, addressed in later sections under this and other elements.

6.2.2.4 Further development of the principle of motivation: external v. relational insightfulness

To sustain the envisioned progress, some observations about the pros and cons of explicating and furthering principles through codification are of additional interest. Related to the discussions at the start of the section, the need for realistic expectations with regard to the principled engagement of administrative decision makers and explainers is of relevance. The chapter identified the formation (in administrative legal scholarship) of an unhelpful dichotomy between notions of formality, legality, legitimacy, inflexibility and in-humaneness on the one hand, and informality, flexibility, creativity, humaneness, and value sensitivity on the other. A more principled engagement was seen to require more ‘free’ discretionary space, flexibility, and therewith informality, which then raised concerns about how this could invite arbitrariness. This could be read as an argument *for* codification. But there were also concerns about how explicating a principle through codification could ‘cut it off’ from further societal development, and others argued that fundamental protection from Administrative power abuse needs to reside outside the codified Administrative legal system itself. In other words, principles need ‘fresh air.’

That may be, but the chapter’s research critically supported the thesis’s argument for explicit and publicly known norms for proper explainer behavior and insightful testimonial processes. The ‘light’ codification of explanation in this domain was shown to be problematic in this respect: decision subjects have more extensive rights on the basis of the principle of motivation (+ due process) than what can be gleaned from the actual testimonial rules, but can’t be expected to know so themselves. In addition, the absence of a bespoke regime for evidence does not work out in their favor when they don’t know what to barter for, and law itself does not provide them with sufficient clues. In light of this, the acknowledged ‘external insight’ rationale for giving reasons (i.e. a statement of reasons needs to allow external parties assess compliance with the much richer principle of motivation, of due process, proportionality, and other applicable principles) is not what needs more catering to: the principles need to be made to count for the relation that they are most salient for, which is that of civil servants and explainees. The further, societal development of the principles needs feedback from the ground, rather than from the expert circles that have thus far not progressed the insightfulness of administrative practices.

6.2.2.5 *The focus on meaningful reviewers, problematized (introduction)*

In the GDPR's ADM regime and other ADM explanation governance efforts, the human explainer is fore fronted as meaningful intervener, a person with the obligation to engage with decision subjects' problems and objections, and who has the know-how and authority to 'fix' whatever went wrong on the basis of the review investigation. The kind of knowledge, authority, and capabilities that are expected of this person have also been advocated for (for many decades) in scholarship on the Administrative domain for 'normal' review decision making. It is promoted that reviewers engage with the goals and affordances of different laws and policies, i.e. to explore the larger administrative landscape, and make creative use of their discretionary space. They are asked to investigate review subjects' *actual* problem so that the (possibly unhelpful) eligibility slot through which subjects introduced their need to the system does not restrict the solution space. The State itself, too promotes tailor-made review processes. The chapter discussed experiments of so-called informal approaches, and some observations about these are cited under later sections.

The point to make here is that these ideals of 'meaningful intervention' sketch an approach to decision making that (see above) is not always promoted, and even frustrated, in ground rules, policies and procedures. Where this already does not stimulate the kind of moral explainer engagement that is called for in ADM times, the focus on review itself is arguably problematic, too. This argument is introduced here, to be returned to under other elements.

For element one, the main issue to note is how the Awb's review approach risks to 'unhinge' the type of caring explainer disposition that the ADM efforts advocate. Administrative bodies are allowed several instances to repair reasons for decisions, which would be less problematic if the same reparation possibilities were afforded to decision subjects. The reviews themselves are done by employees with more legal training. The chapter noted this as a premeditated reduction of the knowledgeability of initial administrative decision makers, and speculated on an additional problematic consequence: since the decisions of the less legally trained primary decision makers stand to be corrected by employees that (again, legally) outrank them, this could arguably *dis*-encourage them from the type of critical investigative engagement that element one requires, and even stimulate to 'simply' follow procedure.

More in general, the domain research supports scholarly critique on how locating the most salient 'rights' against the 'wrongs' of ADM practices on review level relies on the (mental, temporal, emotional, financial) ability of decision subjects to enter into review procedures.⁹⁸⁹ Ironically, 'pre-emptive' solutions are not necessarily helpful here. E.g., as was discussed above, using automation to identify and select types of decision subjects that can be expected to need to rely on review because the system

989 Binns, 'Human Judgment in Algorithmic Loops'.

that is designed can't fairly treat their types of situations well, are incurred with additional human investigative steps while others aren't, which still gives the latter unfair advantages.⁹⁹⁰

6.2.3 Observations from the GP domain

6.2.3.1 *Bad knowledge practices flourish in absence of explanation obligations and relations (it should not need repeating)*

Ethical principles of the medical domain are much named as inspiration for 'AI ethics',⁹⁹¹ and a Hippocratic oath for mathematicians has even been called for.⁹⁹² But self-regulation alone is not enough. Worse, it undermines progress. The domain research testifies to how the historical absence of legal obligations to engage in explanatory and justificatory exchanges with patients had a profound negative impact on the quality of medical knowledge and practices. Medical knowledge (and physicians) lacked crucial information and understanding. Strong hierarchical cultures combined with a lack of justificatory traditions sustained exploitative practices, and a cult of silence around medical controversies whose ontologies failed to inform progressive development as a consequence. Harmful ideologies flourished in such environments. The pattern was and is predictable: currently, the 'colonial' practices of commercial AI developers (such as, training on data from populations in less legally protected regions and experimenting with models in such environments, the value extraction enabled this way, the disregard for and exclusion of knowledge makers, practices and infrastructures of affected populations) are being called out and juxtaposed by different natured initiatives.⁹⁹³

The thus produced lack of 'meaningful knowledge relations' with patients proved to be an obstacle around the paradigm shift to informed consent, and later to SDM. Neither party was well prepared for more honest and equal social-epistemic relations. GPs were concerned that under such conditions, it is irresponsible to assume that patients have sufficient medical self-understanding to responsibly participate in decision making. And they were right—but GP's own insight into how to serve patients information-wise

990 Binns and Veale, 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR', 329.

991 A critique on doing this naively was cited earlier; The AI and Robotics group at the Tilburg Institute for Law, Technology and Society, 'Response on the draft ethical guidelines for trustworthy AI produced by the European Commission's High-Level Expert Group on Artificial Intelligence.', 31 January 2019.

992 Sample, 'Maths and Tech Specialists Need Hippocratic Oath, Says Academic'.

993 See for example DAIR's statement about "Community, not exploitation" 'The DAIR Institute', last consulted 20 November 2022, <https://www.dair-institute.org/about>; Linnet Taylor, 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally', *Big Data & Society* 4, nr. 2 (1 December 2017); and among others Irani et al, 'Postcolonial Computing'; Shakir Mohamed, Marie-Therese Png, and William Isaac, 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence', *Philosophy & Technology*, 12 July 2020; Balayn and Gürses, 'Beyond-Debiasing: Regulating AI and its inequalities'.

was lacking too. And so, it seems, was their motivation to change this. Repeated legal evaluations revealed a gross lack of compliance with informed consent obligations.

For AI-infused times, such histories are an obvious lesson. Especially in high-tech environments, which medicine certainly became, the obligatory relation between what is easily separated out into ‘doctor side expertise’ and ‘patient side information’ is pertinent. The point is to work towards explainable trustworthiness. As the introductory chapters discussed, that aim is actively undermined by those who downplay the value of understanding, or the ability of humans to provide the necessary reflective insight in the first place.⁹⁹⁴

With that in mind, an Algorithmic Impact Assessment (AIA) process trial report from the medical domain is interesting to cite. The Ada Lovelace institute trialed an AIA process for Public Health procurement of private sector decision support systems.⁹⁹⁵ In line with the common understanding of the AIA as an instrument to provide evidence of meeting accountability needs, and following Wieringa, they build on Bovens’ conceptualization of accountability: obliging an actor to explain and justify their actions to a forum.⁹⁹⁶ They further specified such a forum must have “the capacity to deliberate on the actor’s actions, ask questions, pass judgment and enforce sanctions if necessary.”⁹⁹⁷ What they *add* to common instructions for AIA processes is a separate goal called “reflection/reflexivity.” The exercise demands a critical dialogue between the developers and affected individuals (“clinicians, patients, and society”) about any pro’s, cons, and ethical concerns around the design, development, and envisioned use of the system. It includes “examining or responding to one’s – or that of a teams’ own practices, motives and beliefs during a research process,” including individual biases “and ways of viewing and understanding the world.” It obliges to engage with domain-specific considerations as well as those identified in ‘algorithmic literature’: discrimination and marginalization (including the exacerbation of existing health inequalities), interpretability issues, data justice issues, effects on the doc-patient relationships, and so on. The applicant team needs to narrate best and worst-case scenarios,⁹⁹⁸ and specifically engage with the perceived seriousness and ‘difficulty and detectability’ of imagined harms.

994 See for example section 2.2.4

995 Applicant tech-designers are ‘processed’ before they get access to the NHS image database they need to train their model. ‘Algorithmic Impact Assessment’.

996 Mark Bovens, ‘Analysing and Assessing Accountability: A Conceptual Framework1’, *European Law Journal* 13, nr. 4 (2007): 447–68; Wieringa, ‘What to account for when accounting for algorithms’.

997 Metcalf et al, ‘Algorithmic Impact Assessments and Accountability’.

998 Amore too suggests that this is a useful exercise, inspired by how Feinman asked the Challenger launch team to explicate their piece of the risk-prediction puzzle: what it would mean for their ‘bit’ (instrument, knowledge) to fail. Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*, 147.

This reflective exercise starts the AIA, and an unsatisfactory engagement terminates a team's application process.⁹⁹⁹ Later on in the AIA process, a participatory workshop ensures additional engagement of the applicant knowledge makers with doctors, patients, and other interested parties. The Institute recommends to include diverse and underrepresented groups, representatives of such groups, and 'critical friends' (knowledgeable on AI issues and bio-medicine) in the team, and to ensure the social safety of the setting. Last, but in light of the domain-based concerns, certainly not least, they also advise to adopt the AIA under a legal instrument to boost its realistic power.¹⁰⁰⁰

The institute, in other words, promotes the creation of a non-optional knowledge tool. They promote to expose typically (and problematically) separated parties to each other's expertise, to redistribute social-epistemic authority, to infuse the process with critical wisdom from outside of the development circles, and to support the process with multidisciplinary guidance.¹⁰⁰¹ The documentation of such a process could support doctor-patient interactions in later stages when the tool is used, helping doctors to help their patients to invest their trust responsibly.

6.2.3.2 *The technological complexity argument*

The lack of obligations to justify knowledge claims to patients also supported the cultivation of *mere* authority as an acceptable (even beneficial) characteristic of relations with large information inequalities. An irresponsible narrative about the conditions under which 'patient trust' establishes was conceived. The (justifiable, under terms) claim that there is a therapeutical dimension to patients' trust in their more knowledgeable doctors was perverted into a problematic claim that patients necessarily trust their doctors blindly, because medical knowledge is too complex for them and they can't meaningfully understand it.¹⁰⁰²

The same argument is used today to water down concerns about the lack of understanding of subjects of AI-supported decisions. It therefore merits to compare Katz's 1984 contestations of the false premises that underlay the argument with synchronous arguments in the current discourse. The original arguments are printed with emphasis. First of all, the technological complexity argument was said to obscure other reasons that medical experts had to avoid justifying their understanding. These included *blind loyalty to their elders* (alive in accounts of e.g. Broussard and Lepore of the influence of AI's original brat pack's status and ideas¹⁰⁰³); *a lack of own understanding* (which is increasingly established to be the

999 'Algorithmic Impact Assessment', 46 An elaborate documentation scheme is also embedded in the AIA, which will be further discussed under element 4.

1000 'Algorithmic Impact Assessment', 68.

1001 'Algorithmic Impact Assessment', 58–61.

1002 Section 5.2.3.3

1003 Broussard, *Artificial Unintelligence*; Lepore, *If Then*.

case in AI: in terms of mathematical complexity, causality, and especially in terms of normative dimensions¹⁰⁰⁴); *paternalistic convictions* (recognizable in arguments such as that technological innovators will save the day if you let them¹⁰⁰⁵); and *a failure to recognize the humanity of certain groups of patients* (resembling the automated handling, marginalization and precarization of groups today¹⁰⁰⁶). In addition, the premise that ‘*medical knowledge*’ is a separate, clearly describable thing was problematized (see Amore on the contingency of AI with the world it exists in¹⁰⁰⁷), as was the notion that even in absence of explanation conversations, *doctors always understand their patients’ health states well enough* (among critiques on the AI field is the fact that we have let a highly homogeneous group of persons develop data-driven knowledge about others under no obligation to understand how subjects are affected socially, for which understanding they also lack the necessary investigative skills and capacity¹⁰⁰⁸). And then there were the problematic narratives that *patients’ meaningful understanding would depend on their individual knowledge of medical technicalities, which is impossible* (simply replace ‘medical’ with ‘algorithmic’ or ‘AI’); *that doctors are generally trusted in the first place* (rather than overpowered, which is equally problematic in the digital age), *and that when they are trusted, they are trusted on the basis of their knowledge* (in absence of meaningful information positions, explainees have no choice but to trust the social status of decision makers.) The last assumption that was named was that *knowledge and decision making about patients can proceed responsibly without patients’ meaningful understanding of, and informed participation in it*. The untruth of this was defended in the thesis’s introductory chapters. The statement is recognizable in a frequently defended dichotomy which holds that ‘explainability’ comes at the detriment of ‘accuracy’ of algorithmic systems, which is said to result precisely from ML’s inscrutable ways.¹⁰⁰⁹

1004 See for example Malik, ‘A Hierarchy of Limitations in Machine Learning’; Synced, ‘Yann LeCun Quits Twitter Amid Acrimonious Exchanges on AI Bias’.

1005 To cite just three moderate results after a google search on can AI save the world / us: <https://aiforgood.itu.int/8-ways-ai-can-help-save-the-planet/>, <https://www.forbes.com/sites/neilsahota/2020/07/27/will-ai-save-us-from-ourselves/>, <https://medium.com/@LanceUlanoff/deepminds-stunning-breakthrough-shows-how-ai-could-save-us-30b360845cf1>

1006 As was cited at various points throughout the thesis.

1007 Amore, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*.

1008 To quote Broussard, ‘To recap: we have a small, elite group of men who tend to overestimate their mathematical abilities, who have systematically excluded women and people of color in favor of machines for centuries, who tend to want to make science fiction real, who have little regard for social convention, who don’t believe that social norms or rules apply to them, who have unused piles of government money sitting around and who have adopted the ideological rhetoric of far-right libertarian anarcho-capitalists. What could possibly go wrong?’ Broussard, *Artificial Unintelligence*, 85.

1009 Section 3.2.4.1

The appeal of the technological complexity argument continues to be very strong. As an excellent conduit of wrongful knowledge making, it needs to be engaged with in any effort towards explanation regulation. It should certainly inform considerations about to what extent to let a field of expertise self-regulate. This is addressed below.

6.2.3.3 *What's keeping law from 'realizing the best possible care'?*

In the small set of contemporary professional / ethical standards and education literature that was studied, and in GPs' own scholarly contributions, the objectives of element one are strongly represented. Identified were, *among other things*, obligations to be(come) critical, multi-disciplinary knowers, learners, and contributors to medical knowledge and public health policy; to acknowledge the value of doubt and responsible distrust; to fend for qualitative SDM methods in the face of (the push for) quantified standardization, to be alert about the existence of unfair medicine and one's own biases, to be aware of the instability of diagnostic concepts and the performative effects of sharing them with patients.

All these notions are absent in the WGBO's 'information obligations.' But per the Handbook of Dutch Health Law, the *standards* need further development. Doctors were said to need more guidance in face of fundamental questions raised by *new* types of medical knowledge (think: AI-driven predictive diagnostics, genetics). The weakness of the 'novelty' argument will not be re-engaged with here, it was reiterated in this chapter's introduction. The point to make here is that, without a clear legal anchor for this need, it will be hard to force compliance with the envisioned standard. The recent update of the WGBO would arguably not suffice. As said, the obligations don't address these issues. The law also avoids to use the word that produces the purported necessary conversations in the first place: 'explanation.' Neither does the law oblige to check for patient understanding, another compliance-supportive obligation, one that was part of professional standards for a long time already.

The *need* to force compliance can't be predicted for the future, but past experience teaches to be at least conservative with regard to positive expectations. The domain analysis showed how the critical understanding of medical knowledge, or any understanding was not naturally shared with patients. This was true when informed consent was based on more general constitutional grounds, after enactment of the WGBO, and more than a decade after it – and after an implementation process engaged with the compliance problems.

The WGBO therewith expresses a weak institutional promotion of accepted standards in a field in which power and information inequalities are very large; a field whose practice spheres and cultures are historically explanation and justification-averse. A choice that begs to be defended, or so the chapter argued. Legemaate argued that the legal aim of realizing the best possible care itself needs 'perpetual care.' Critical

development needs space, but leaving too much interpretation up to the field can fail to force behavioral change of values that express stronger in society than in the medical field.¹⁰¹⁰ The question is, do they express in Lawmakers' circles? Their lack of address leaves an important function of explanation rules unused: to 'demonstrably aim' for values that need further promotion and support *in* society—also to support those employed in explainer positions. When doctors are legally obliged to discuss questions of knowledge, they have more power to demand explainability of the 'new' knowledge they can, by now, not avoid to work with. AI is being implemented at all possible levels of medical technological development. Element one's 'obligation to protest' is arguably of use, as well. As Pierce et al. argue, the 'automation' of what frequently are already unstable or flawed diagnostic concepts, whose responsible use and further development requires deep engagement from human doctors, complicates their moral responsibility for patient care.¹⁰¹¹ In addition, ensuring that *patients* are aware of the knowledge-related obligations of doctors, by lifting the obligations out of professional standards and creating a kind of shared ownership, would enable both sides to 'team up' and make a stronger claim.¹⁰¹²

6.2.4 Suggested emphasis for element one

What stands out from both domains' research is the need to address explainers who are untrained, uneducated, or—worse—reluctant to educate themselves on characteristics of epistemic oppression. Even if that is not the case, their embedding in decision practices with their own history, hierarchies, and traditions may impede their investigative tendencies, and/or their inclination to protest when they do have reason to. This is more problematic in the face of technological push from fields that historically cannot be trusted about their own intentions, knowledgeability, and sensitivity.¹⁰¹³

To address these issues, any contextual application of the Model would need to be more specific in what it asks to investigate: not simply speak of 'oppressive knowledge practices' but explicate the types of epistemic wrongs, harms, and dynamics to look out for. Perhaps Element one's general description should already include some words on investigating for potential gaps in decision makers' knowledge. The idea, as explained in Chapter 3, is to make use of legal rules' prescriptive potential, in other words, to rely on them to *make* this happen. This will be picked up in the last section of the chapter.

¹⁰¹⁰Section 5.3.2.1

¹⁰¹¹Pierce, Sterckx, and Van Biesen, 'A Riddle, Wrapped in a Mystery, inside an Enigma'.

¹⁰¹²Shunryu Colin Garvey, 'Unsavory Medicine for Technological Civilization: Introducing "Artificial Intelligence & Its Discontents"', *Interdisciplinary Science Reviews* 46, nr. 1–2 (3 April 2021): 7.

¹⁰¹³Section 6.2.3.2

6.3 Engaging with the social-epistemic positions of explainees (element two)

6.3.1 General observation

The second element represents what can be described as the second preparatory step, or phase, of ‘explanation due process.’ The obligation here is to investigate whether explainees’ social-informational positions are, and were, of sufficient quality to participate responsibly and meaningfully in the decision-and explanation processes. The obligation pertains to individual as well as group levels, and again the effort needs to be demonstrated to explainees.

The objectives of element two are at best implicit in the studied legal explanation rules of both domains. Both paradigms were found to be criticized for their unfounded assumptions of legal, bureaucratic, and health literacy, their assumptions of subjects’ self-sufficiency, their focus on individual autonomy. In other words, the ‘care’ for responsible participation that ADM explanation efforts mean to ‘reinstate’ is not a natural resource of existing explanation paradigms, and should not be expected to flourish without clear instruction. Put differently yet again, even when AI methods would be (made) ‘explainable,’ there is no guarantee that decision subjects will benefit in terms of element two. At a time when stages of human discretion and automation are increasingly hard to pull apart, the discrepancy between domain-specific governance and the more engaged explanation regimes for ADM are increasingly unproductive and prevent a truly progressive update that relates existing needs with those that are indeed new(er).

6.3.2 Illustrations from the administrative domain

6.3.2.1 Does the “European unease” with automation need a progressive update?

Support for element two was clear in older and contemporary calls to improve Administrative decision makers’ engagement with decision subjects’ social and informational situatedness. Critique pertained to a variety of causes. Among them misplaced assumptions about citizens’ ‘bureaucratic’ and legal literacy and informational powers, wrongful (as in: unfounded and unjustified in the face of adverse scientific knowledge) ‘model citizen’ design in Administrative laws, a lack of protection for citizens against nasty bureaucratic processes, and reduced meaningfulness of participation due to the persistent tendencies of administrative bodies to reduce their explanation burdens. To follow up on the previous element’s discussion on insufficient uptake of such critiques, an observation here is that concerns about the exacerbating role and influence of automation and ‘digitalization’ of the public domain can only be well-placed when they are indeed understood as exacerbations.

Put differently, what we have started to ask for in AI-infused and simpler ADM contexts will need to be asked of the ‘analog’ context if the Administrative principles of motivation and due process are to be of *meaningful* use in ADM contexts, as is being argued. Coming at it from the other, digital side, Binns and Veale’s observations support the argument. They consider that “blurring [the] bright line” between automated and human decision making in the GDPR may prove controversial in light of the purported “European unease” with automation,¹⁰¹⁴ but warn against regulation whose strongest protection against ADM is triggered by unrealistic assumptions of where ‘human scrutiny’ exists, and what can be expected of it. The concern that this allows creative decision design(ers) to escape legal protections was related to the Awb’s restriction on what ‘counts’ as decision-material that needs justification. The argument to make here is related, but broader: an un-carefully executed ADM-protection regime risks to add to the list of wrongs and harms that happen undetected.¹⁰¹⁵ In such cases, the underlying regime has more, not less, work to do. When that doesn’t ‘help,’ and ADM-regulation is seen to become too much hard work, the authors warn for judicial tendencies to shift the focus to ex-post oversight: to focus on the consequences, not the causes of ill treatment. They name recent ‘gymnastics’ of the Court of Justice of the EU who has the last word on the interpretation of EU Data Protection laws. The court might seek to “transform stubborn ex ante concepts like lawful bases into ex post oversight.” This would severely weaken the burden on ‘first explainers’ that this thesis seeks to strengthen.

6.3.2.2 *Don’t ignore the messenger: lessons to learn from the ‘analog’ Administrative domain*

It is currently unclear whether the further development of the analog and digital governance of the domain will be addressed in an integrated way by the Dutch lawmaker. Below, different arguments are discussed and related to findings from the domain study.

The chapter briefly discussed the ‘spat’ between the Council of State and the Legislator about a broad exception clause in the Dutch GDPR implementation law.¹⁰¹⁶ The clause establishes a general legal ground for ‘simple’ administrative ADM, i.e. in absence of ‘modern methods of profiling’ that risk to hold ‘negative features of a certain group (...) against an individual who does not possess it’ (which the chapter pointed out as

¹⁰¹⁴Binns and Veale, ‘Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR’.

¹⁰¹⁵“the potential for selective automation on subsets of data subjects despite generally adequate human input; the ambiguity around where to locate the decision itself; whether ‘significance’ should be interpreted in terms of any ‘potential’ effects or only selectively in terms of ‘realised’ effects; the potential for upstream automation processes to foreclose downstream outcomes despite human input; and finally, that a focus on the final step may distract from the status and importance of upstream processes” Binns and Veale, 332.

¹⁰¹⁶Section 4.2.3.3

a painfully inadequate description.) The Council had advised against it, citing how even ‘simple automation’ had thus far proven to be a source of harm for large groups of citizens. Among other things the State deferred to administrative bodies’ own rule making powers, arguing they had failed to design adequate discretionary space in their executive policies. The argument assumes two types of neutrality where there is reason to do the opposite. One, Administrative automation referred to, like systems before and after it, were built to do exactly what they did, serving the (detrimental) spirit of the underlying law and policy, second, it was well known that more discretionary space for administrative bodies does not naturally lead to more attention for responsible participation possibilities for citizens—even the opposite, as the WMO cases showed.¹⁰¹⁷

The clause stood. But where the Council around that time argued to embrace article 22 and honor, develop and explicate the principles of due process and motivation to make it work, a more recent set of recommendations for ‘digitalization and legislation’ professes doubt that the current legal landscape is adequate. “Maybe,” the Council considers, “we need to acknowledge that the two worlds are hard to unite.” They consider the take-up of an additional ADM regime *in* Administrative law: new rules, new principles even, including a ‘right to information’ and to ‘meaningful explanation.’¹⁰¹⁸ With that, the Council risks to posit itself *en route* to unhelpful automation-dependent constructs comparable to those sketched above: problems that the State ironically prevented from establishing by sidelining article 22. Other suggested solutions, too, may not be very useful if lessons about the current regime are not taken on board. Some examples:

- The Council writes how “choices and estimates” made by “ICT professionals” in their automation (“digital translation”) of law & policy should be made understandably available to “citizens, their (legal) representatives and others,” so that they can check whether this is done in a “good, well-considered and careful” way.¹⁰¹⁹ But the domain study showed how realistic citizen scrutiny of complex administrative laws, let alone law-to-policy translations, is easily illusionary in analog contexts already. And even challenging for legal representatives.
- Taking subjects seriously in terms of element two means understanding how they are positioned in the much larger web of administrative decision practices. Patterns of marginalization appear through insight across different decisional contexts, but the Awb wasn’t built for such labor. Again, when more human discretion is envisioned, which is to be expected in light of calls for ‘humane, meaningful intervention,’ then this needs attention. Interestingly, the need to grapple with the interplay of different

¹⁰¹⁷Sections 4.2.3.5, 4.2.3.6, 4.3.3.5

¹⁰¹⁸‘Digitalisering: wetgeving en bestuursrechtspraak’ (Raad van State, May 2021), 75.

¹⁰¹⁹‘Digitalisering: wetgeving en bestuursrechtspraak’, 20.

laws and how these lead to “unintended” hardship cases is included as one of the ‘headache accounts’ in the Annual Report of the Judiciary.¹⁰²⁰

- More fundamentally, the Council ultimately relies on “real people” to administer real individual justice in ADM contexts.. in review.¹⁰²¹ This unfairly burdens subjects with review procedures, and also relies on the individually focused ‘disproportionate hardship’ regime. Earlier voiced concerns apply and will not be repeated here. To add is how the domain study revealed various obstacles for responsible explainee participation such as inscrutability of information creation by administrative bodies, challenges to get this corrected, unclear regimes and unexplicated principles for what may count as facts, interests, evidence. These complicate a responsible *start* to procedures, and such starts are not necessarily repairable in review. In the annual advisory reports to the legislator authored by the Association for Administrative Law, Wolswinkel compared the Awb’s regime to digital legislative trends. He points out how the Awb offers little in terms of ‘process transparency’ once a decisional process has started,¹⁰²² and how relevant files are not automatically sent out with a decision’s statement of reasons but upon request. Both things put subjects in bad participation positions per se—but also make it hard for them to decide whether to file for review.
- The Council’s reliance on review procedures as the “excellent” choice for ADM scrutiny, and for meaningful citizen-State interactions arguably plays down the need for ‘first explainers’ to be equipped with the same useful capabilities, and with the access to processes and modalities that prospective reviewers are seen to require. In light of the fundamental concerns about the instrumentalization of ‘bureaucratic armies’ and the reliance on *all* civil servants, arguably, especially those whose positions and tasks risk to undermine their humane, moral engagement, that is a high-risk choice.

6.3.2.3 Integrative approaches

In the cited advice, Wolswinkel concludes that adding ‘process transparency’ obligations in the Awb would let decision subjects participate more meaningfully. He argues these could be derived from the principle of diligent preparation. The Awb’s process transparency regime for public procurement procedures is already more elaborate and could serve as inspiration too. E.g., a redesign of case file management into a ‘dynamic’ system that is accessible (and understandable) for decision subjects also helps to prepare parties for a meaningful explanation process.¹⁰²³

1020 ‘Jaarverslag van de Rechtspraak’ (Raad voor de Rechtspraak, May 2022), 22.

1021 ‘Digitalisering: wetgeving en bestuursrechtspraak’, 80.

1022 C.J. Wolswinkel, ‘Transparantie en openbaarheid: preadviezen 2022’ (VAR, 2022), 189.

1023 Wolswinkel, 201.

Taking the need for ‘meaningful participation’ obligations further, Ranchordás argues for the legal instrumentalization of ‘administrative empathy,’ which she defines as “the ability to acknowledge, respond, and understand the situation of others, including their challenges and concerns.” Such empathy should *alert* decision makers to their “moral choice[-making],” and *require* them “to understand citizens’ needs in the context of government transactions and regulations.” The concept, she argues, should be operationalized into specific obligations to avoid a moot “re-humanization” of automated Administration (especially of the welfare domain), “at a time when empirical evidence suggest that humans are becoming less emphatic than previous generations and humans-in-the-loop do not take meaningful actions.”¹⁰²⁴ Bureaucracy needs a redesign, and legal instrumentalizations of Administrative Empathy can be used to *expose* externalities of automation on vulnerable subjects,¹⁰²⁵ and *repair* their capabilities, and with this, their rights, to participate safely and responsibly. She for example suggests to take up a duty to “forgive” decision subjects for procedural non-compliance that follows from vulnerability, including such bureaucratic overwhelming as is frequently described in critiques on automated welfare states. As such the proposal is a direct engagement with the above-described inadequacy of the ‘disproportionate hardship’ standard. The main difference between Ranchordás’ argument and responsible participation in terms of this thesis, is the focus on where participatory vulnerability establishes. Where the Model focuses on (intentional and implicit) vulnerabilizing affordances of law & policy, the focus of Ranchordás’ article is on additional debilitating effects of (especially) automated/digitalized policy effectuations.¹⁰²⁶

6.3.2.4 Using AIAs to foster explanatory clues for responsible participation

Kaminski and Malgieri consider the value of using the GDPR’s Data Protection Impact Assessment (DPIA) as a tool to (also) foster clues for explanation. This way, the individual rights-based protections of the GDPR can be (more) usefully related to its system-level governance: the level where those kinds of potential harms are addressed that don’t express individually and escape the GDPR’s (and other regimes’) individually oriented safeguards.¹⁰²⁷ In (what they envision to be) continuous auditing processes, they consider how DPIA’s produce ‘webs of explanation’: documented interactions between parties to impact processes. To make their interactions meaningful, developers, members of internal and external boards and bodies of oversight and assessment, and other assessment parties need to make their knowledge and information understandable for each other: they need to mediate their

¹⁰²⁴Ranchordás, ‘Empathy in the Digital Administrative State’, 11.

¹⁰²⁵I.e. how they are punished, literally or in terms of missed chances, for innocent mistakes, for bureaucratic overwhelmedness, for not knowing their rights and policy affordances, for a lack of ‘digital literacy,’ et cetera.

¹⁰²⁶Ranchordás, ‘Empathy in the Digital Administrative State’.

¹⁰²⁷Margot E. Kaminski and Gianclaudio Malgieri, ‘Algorithmic impact assessments under the GDPR: producing multi-layered explanations’, *International Data Privacy Law* 11, nr. 2 (2021).

information.¹⁰²⁸ Those mediations can be operationalized to get a grip on what can and needs to be explained to individuals before, during, and after their ‘individual’ decision processes.

As in the earlier cited health care context AIA example, the authors advise to enhance the DPIA with an interdisciplinary array of experts, “the involvement of constituents,” and obligatory third-party oversight or assessment. The authors advise how more research is needed, “in particular about how different layers of explanations—systemic explanations, group explanations, and individual explanations—can interact [with] each other.”

The modeled duties of explanation care that this thesis proposes could play a useful role here. The thesis would recommend to regard the impact of individual decisions a system will (help to) make in terms of the larger decisional/institutional context that applies to prospected explainees; to study the interplay of qualitative, quantitative, *and* automated methods at use in the broader decisional process, and to take the decisional domain’s cultural and historical context and justice failures into account. For inspiration on what questions to ask, whose behavior, and which processes to include, and why a mixed-methods approach for such assessments is useful, various studies that were cited earlier provide rich resources.¹⁰²⁹ This section adds Saxena et al.’s study of standardized (and ‘algorithmitized’) Child Welfare risk assessment processes. They show how the human-led, partly algorithmically supported process produces decisions whose ontology is only made insightful through very labor intensive ‘deconstruction’ of all the available materials (e.g. they designed a quantitative method to analyze conversation transcripts, qualitatively investigated the standardized and automated elements, related the two to understand their interaction.)¹⁰³⁰ Such studies, or education materials based on them, could provide ‘explanation clues’ on and for all levels—and certainly deserve a place in the education and training of decision makers in such contexts (see element one).

¹⁰²⁸Kaminski and Malgieri, 142.

¹⁰²⁹Among which, Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*; Balayn and Gürses, ‘Beyond-Debiasing: Regulating AI and its inequalities’; Amoores, *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*.

¹⁰³⁰Devansh Saxena et al, ‘How to Train a (Bad) Algorithmic Caseworker: A Quantitative Deconstruction of Risk Assessments in Child Welfare’, in *CHI Conference on Human Factors in Computing Systems Extended Abstracts* (CHI ’22: CHI Conference on Human Factors in Computing Systems, New Orleans LA USA: ACM, 2022), 1–7.

6.3.3 Observations from the GP domain

6.3.3.1 *Times are different now?*

As was also engaged with in element one, The Handbook of Dutch Health Law called attention to new and fundamental questions raised by new medical knowledge practices (notably, predictive diagnostics) and how that poses challenges for Shared Decision Making. A notion also found in discourse from the medical field itself. But responsible patient participation in relation to novel knowledge practices is only meaningfully pursued with a proper understanding of how ‘times are different now’—and how they aren’t. The field of genetics/genomics is a case in point: its very troubled history¹⁰³¹ warns to be very careful about what research to pursue, and to take all possible precautions to ensure that a responsible medical practice is established. It is also a field that AI developers saliently center their attention on, producing many of the novel questions the Handbook refers to.

One would therefore hope that (medical and other) professionals active in these fields are (already) alert; also with regard to make sure that decision subjects (such as those who participate in genetic research) are socio-epistemically positioned so as not to be coerced. The GP context exists ‘downstream’ of what happens on these R&D levels, and in that sense is dependent on the governance of these levels. They cannot realistically prevent (the effects of) a lack of alertness there, but they do have a responsibility towards patients who are interested to undergo novel diagnostics and treatment.

The section leaves GP’s ‘downstream’ context for a moment to consider this gap. In a 2022 interview, the Dutch Minister of Science and Education—politically a new recruit, entering with an international star status in the natural sciences—expressed what seems an alarming lack of awareness of the ‘politics’ of medicine.¹⁰³² Referring to Dutch researchers’ involvement in controversial Chinese Uighur DNA research,¹⁰³³ the Minister argued that this was a ‘new problem’ to deal with: “fifty years ago, we did not think of genetic information in terms of human rights.” In the same interview, the Minister expressed his surprise about what he saw as a concerning ‘politicized’ societal response to the outgoing government’s ‘scientifically grounded’ pandemic policy measures. The response inspired (his) concerns for a future in which governments need to be able to use novel scientific insight to tackle other complex problems, too, such as those of climate change.

¹⁰³¹As was discussed in Chapter 3, see section 3.2.3.1

¹⁰³²Lucas Brouwers and Patricia Veldhuis, ‘Robbert Dijkgraaf: “Politici hebben heel besliste meningen. Daar moet ik aan wennen”’, *NRC*, 11 March 2022, <https://www.nrc.nl/nieuws/2022/03/11/robbert-dijkgraaf-politici-hebben-heel-besliste-meningen-daar-moet-ik-aan-wennen-a4100635>.

¹⁰³³Elmer Smaling, ‘Controversial DNA Testing? Address the Ethical Issues’, *Erasmus Magazine* (blog), 14 October 2021, <https://www.erasmusmagazine.nl/en/2021/10/14/controversial-dna-testing-address-the-ethical-issues/>.

The remarks seem to express a concerning lack of understanding of the political dimensions of both science and (ironically) policy; of what it means (or should mean) to justify policy choices, and therewith to engage with citizens' and research subjects' social-epistemic situatedness.¹⁰³⁴ With such notions on lawmaker level, it is perhaps not surprising that Dutch health law abstracts from the political dimension of medicine in its informed consent obligations – and all the more pertinent to anchor the more responsible, inclusive explanation practices in law. This would support the well-informed participation, i.e. the *responsible* trust of subjects in all contexts in which science is a prominent ground of decisional practices.

6.3.3.2 Professional norms of engagement to explicate, boost, & codify

In the studied non-legal materials from the domain, support for element two's objective of tending to explainees social-epistemic situatedness to enable meaningful subject participation found ample expression, also in explicit obligations. The need to make up for a lack of patient-side medical knowledge after centuries of paternalistic and otherwise authoritarian practices was acknowledged. Among contemporary norms were the need for medical explainers to understand their practice from patients' point of view; the instruction for GPs to chart their patient populations' social-epistemic positions and how these inform their informational needs; the general obligation for medical practitioners to verify patients' understanding. Scholarly (including GP's) literature from the domain discussed how patients cannot be expected to mention an important value, fact, or consideration if they don't know that it has medical relevance, and that responsible SDM requires qualitative capabilities to support that this happens. It was also emphasized that GPs have a role to play in critically understanding the larger public health system: as such they have raised their voices against law & policy makers' unrealistic assumptions of a highly autonomous, 'self-sufficient' citizen-patient population. They are relied on to be alert with regard to trends of medicalization, and also criticized when they are seen as instrumental in such trends (e.g. by prescribing ADHD medicine on a broad scale).

But in a world where medical practices are still known to suffer from very many historical and contemporary discriminatory wrongs, where discriminatory GP diagnostics are an urgent point on the Dutch Capital's 2022 Public Health agenda, such norms deserve to be anchored in a way that they translate to legal obligations and corresponding patient rights more directly. Patients need to know that they can expect that their concerns and inclinations 'to invest their trust or not'¹⁰³⁵ in this regard are engaged with and are a valid point of discussion – and doctors need to know that they

1034 Aviva de Groot et al, 'Technologie is niet neutraal, dat zou Dijkgraaf moeten weten', *ScienceGuide* (blog), 5 April 2022, <https://www.scienceguide.nl/2022/04/technologie-is-niet-neutraal-dat-zou-dijkgraaf-moeten-weten/>.

1035 O'Neill, 'Accountability, Trust and Informed Consent in Medical Practice and Research'.

know. To reiterate, this is a context in which meaningful justificatory conversations with patients don't tend to establish *without* strong legal intervention.

Currently, the WGBO's provisions don't explicitly engage with patients social-epistemic information positions at all, especially not those that exist on group rather than individual levels. Patients' informational needs are not treated as a given, but measured against what would be directly useful to know for an individual 'reasonable person' who is about to make a choice. The law's horizontal, 'contractual' governance of what is a very unequal social-epistemic relationship in reality was criticized as a bad fit both from within and around professional circles.¹⁰³⁶ An interesting suggestion from Legemaate was to introduce 'patient buddies' to strengthen patients positions in consultation: a suggestion that was also made for the bureaucratic welfare context by Vredenburg.¹⁰³⁷ This point will be repeated in the next element, as it is of special relevance there. For here, the lack of legal address is mainly presented as another example of non-novelty and the need to understand where problems come from. The meaningful participation of decision subjects is not secured in fundamental explanation law in this domain, and 'fixing tech' will not help patients who suffer the consequences of this.

6.4 Practicing interactional justice (element three)

6.4.1 General observation(s)

Under element three, the thesis's take on proper explainer behavior is laid down in demands for explaining, itself. The preparatory phases need to express here: explainees need to be recognized and respected as knowers and rights holders, the social interaction needs to be conducive to those needs, explainers need to actively justify their epistemic authority and promote the right kind of dis/trust: one that is supported by sources from outside of the inevitably unequal power-relationship.

Illustrations from both domains revealed a mostly ungoverned (by law) space when it comes to serving 'interactional justice' in explanation practices. In light of how there is explicit recognition and guidance in professional and ethical norms, legal principles, and research, and in face of evidence that these norms are not sufficiently heeded in practice, law's weak promotion stands out. By not taking a stand, lawmakers express a lack of care to codify the necessary guidance. Explainers are left to their own devices in a time that they are more in need of guidance than ever, and in a time where explainees are in weaker power and information positions themselves. This thesis argues that this situation is not just unproductive but subversive of what our times require. For one, this leaves the development of what *can* be explained up to parties

¹⁰³⁶Section 5.3

¹⁰³⁷Kate Vredenburg, 'The Right to Explanation', *Journal of Political Philosophy* 30, nr. 2 (2022): 224.

who are known to make bad decisions on their own: technological companies. It also leaves professional explainers in too powerful positions vis-à-vis their explainees. Developments around Algorithmic Impact Assessments, again, are interesting to look at, as these increasingly promote to engage *both* parties to the explanation relationship in what can be characterized as ‘knowledge making and awareness raising sessions.’ Such processes, and the documentations they produce, could be used to ‘up the game’ for explanation practices, but only if this aim is made an explicit part of them—and only if lessons from regulated explanation domains are taken on board. One lesson that can arguably be drawn is that the purpose of explanation practices needs to be part of what is explained if respecting explainees not just as knowers, but as rights holders is to be taken seriously.

6.4.2 Observations from the administrative domain

6.4.2.1 (Finally) explicating principles: duties of care as ‘rules of engagement’

Around the Awb’s codification, reports and literature about bad explanation experiences of decision subjects were on the table. These made clear how a meaningful exchange at the explanation stage is dependent on a meaningful, insightful decision trajectory up to that point: an explanation ‘moment’ can’t repair all preceding process inscrutability. A view from the modeled duties of explanation care would argue that ‘reparation’ should indeed not (need to) be the main focus of explanation regulation – but through codification and the practices that followed, this is precisely what *has* become the focus. Moreover, in absence of explicit codification of what quality explanations require, neither the initial explanation stage nor that of review is seen to be served very well. This section runs by several observations that arguably should inform the efforts towards improvement that are being called for, but also need to be known by those who are looking at the regime for inspiration from the perspective of ADM regulation. The next section considers the focus on review specifically.

The ‘principles commission’ (ABAR) report that informed codification already considered the level of review for the elaboration, when necessary, of what they said that honoring the principle of motivation (always) entails: the goals and substance of applicable rules & policy, the choice and weight of any factor and interest that was made to count, external advice, alternative options, and mutual expectations: arguably, the latter obligation for example helps to bring possible participation concerns (in any stage of the process) forward.¹⁰³⁸ But they did not mean for these demands to remain uncoded.

The expectation of the legislating government, expressed at the time, was that all these valid needs and values don’t require explicit codification. And so, the Commission’s elaborate list of items, which recognizes how the principle of motivation requires to testify to a due process, in terms of the Model: that explainers make sure their

¹⁰³⁸ ‘ABAR 1984’, 138–39.

preparatory work expresses in the ‘testimonial exchange,’ remained up in the air. The obligation to provide reasons was codified as an obligation to legitimize an outcome, not to testify to how an administrative body has been an understandable and trustworthy decision partner. This is exacerbated by the fact that the relation between reasons and outcomes is loosened by the ability for administrative bodies to re-reason their conclusions several times. The chapter considered how ‘language is the tool with which to contest the decisional process that a set of reasons represents. For this, reasons need to give sufficient insight. If the argument is of bad quality, this should have consequences, same as arguments *against* a contested decision are only accepted on the basis of their quality.’¹⁰³⁹ These choices are arguably unhelpful in ADM times. Downplaying the value of explicating process over the justification of an outcome opens the door to contemporary (AI-optimistic) arguments that downplay the usefulness of causal understanding.

In combination with several Awb intricacies that were discussed before, this amounts to very weak relations between the principles that are so frequently referred to in tandem in calls for ADM explanation regulation. To reiterate only one intricacy: administrative bodies are obliged to use a designated establishing administrative body’s information about a subject even if the information is known to be incorrect, which asks civil servants to ‘reason away’ important explainee circumstances. The chapter considered how such obligations cannot possibly produce reasoned statements of how unjust knowledge making is avoided.¹⁰⁴⁰ But even when more discretionary space is afforded, the principle fails to express in practice.

Since the Awb’s enactment in 1994, the legislating government’s expectations that the (badly complied with) principle of motivation did not need explicit codification have not been met. Several case illustrations that predate the benefits scandal testify to how administrative bodies have used discretionary space to reduce, rather than enlarge, their motivational burden. Rules of conduct that exist outside of the explanation paradigm, as well, have not been complied with. Administrative complaints procedures, less ‘formal’ spaces than review procedures, were found by the Ombudsman to suffer from a lack of legitimacy. And in trials with review procedures of ‘informal’ character, the lack of insight into what went down and the absence of (re-)reasoned statements made it impossible to check for legitimate testimonial processes to begin with.

The thesis considered how all this puts consecutive Governments *persistent* reliance on administrative bodies’ compliance with the principle of motivation in a suspect light. I.e. the State points the finger at administrative bodies when they don’t comply, but fails to codify the guidance that is obviously needed. It is therefore interesting to see what will become of a Parliamentary motion for the State to develop a principle of ‘meaningful government contact.’ Based on the chapter research, more principles are

1039Section 4.4.4.5

1040Section 4.2.3.1

not what is needed: the explication and codification of the principle of motivation, in combination with that of due process would do. In fact, that effort stands to suffer from a focus on a new principle captured in vaguer terms, that cannot be (expected to be) usefully complied with if explanation guidance remains insufficient.

Claessens was cited to argue for a ‘duty of care’ instead of a principle, which would put the burden of proof on administrative bodies: it will be up to them to prove they have met the described result. It would also put a simple right in the hands of explainees that are currently challenged to fend for what are very scattered and or elusive rights. The thesis agrees, but still argues to bind the required duties, which in fact express in the Model, legally to the Awb’s explanation paradigm. This would serve the Model’s aim of having explicit, public, and enforceable rules that all parties to explanatory exchanges are equally aware of (and at least theoretically have had a say in through the democratic legislative process.) Explainees cannot be respected as knowers and rights holders when their most elaborate rights remain unknown and unreasoned, themselves.

A citation from impact assessment literature supports the ‘shared ownership’ of the rules of engagement that element three aims for. In their report on Algorithmic Impact Assessments, Moss et al. consider the usefulness of AIA’s for how they allow to relate a description of potential and actual harms with “a means for identifying who is responsible for their remedy.” They warn that the success of this depends on the proper address of “social and political power” of a context, and argue to involve affected publics in the identification and description of harms in the first place: publics and individuals *who are not yet aware* of the impact a (type) of system may have on them, nor of their rights with regard to redress, and/or groups that face barriers for getting involved.¹⁰⁴¹

6.4.2.2 *One more time with feeling: pitfalls of the focus on review*

In the Awb paradigm, initial statements of reasons are generally delivered by mail: remotely. Such unmediated delivery arguably requires more, not less substance with regard to what is contained in the statement, but the opposite is true. The review procedure became the main place where elaborate reasons are *made*, as well as the main place for quality control of administrative decisions. Using the initial statement as the lesser modality arguably plays down the value of giving explanations *per se*—hard to understand in light of the principle of motivation’s aim of reducing power and information inequalities, and a questionable choice in an environment that is criticized for its inscrutability for decision subjects. It also puts unjustifiable burdens on explainees. They have little to go on to estimate their need to file for review. And for the type of explainees the thesis focuses on, the burden of going through contestation to get to reasons itself is unfair.

¹⁰⁴¹ Moss et al, ‘Assembling Accountability’, 22.

Furthermore, explainees are confronted with different kinds of explanatory knowledge in review, delivered by a different kind of explainer: ‘the lawyers of the house.’ Both the power and information inequalities grow in such procedures. With reference to the cited consideration for GP domain to make ‘buddies’ available to explainees: in the Administrative domain, an obligatory presence of a ‘buddy’ is an interesting thought. But the level of legal training such a buddy would need to have to provide safe guidance is arguably that of a lawyer; a lawyer with solid principles knowledge. And professional legal help *is* theoretically already available to explainees in the form of subsidized legal aid. That right has however become increasingly moot after a variety of financial and procedural changes to the system over the last decades: for example, Administrative review procedures were no longer supported, a measure installed in the phantom vehicles and Benefits Scandal period. Reinstating that right is the more logical choice.

6.4.3 Observations from the GP domain

6.4.3.1 Be careful what you list for

For AI-infused times, cited discussions in Chapter 5 about the informational exchanges that are/not or should/not be obligatory in the domain’s main explanation law are of interest. In its purported promotion of non-oppressive decision practices, law here inevitably touches upon the need to discuss ‘complex knowledge.’ Below, some considerations with regard to legally requiring ‘items’ of information, types of information, types of conversation, or ‘results’ are presented.

In response to the repeatedly established and grand-scale lack of compliance with informed consent obligations, physicians’ authority with regard to determining the kind of informing that should be done was legally pushed back (a bit) in the domain’s first, bespoke main explanation rules. The ‘items’ that the law named were met with critique by some physicians: explanation is a process, they argued, and cannot be captured in a moment nor a list of things to name. Non-compliance persisted. A legal implementation study followed, and eventually, a major update of the law changed and expanded the list of requirements to do justice to further developments in practice, itself: SDM had established and required a different kind of doctor-patient conversations, conversations aimed at collaborative knowledge and decision making.

The focus in the update was more on curbing authority than on adding additional items of information. Words like informing on ‘necessary’ (treatment) were for example changed into ‘proposed (treatment)’ An added provision speaks of ‘discussion,’ and of inviting questions. The new phrasing is however not entirely devoid of the old ‘doctor knows best’ paternalism. E.g. the new obligation requires a doctor to ‘inform himself (sic) about the situation and needs’ of a patient. Earlier cited critique also pertained to

the lack of *result* the law requires: compare ‘informs the patient in clear language about X’ (currently on the list) with ‘ensures that the patient understands X’ (not required.)

Critiquers of the WGBO’s ‘horizontal’ governance are therewith not satisfied. The law still assumes a more equal social relationship than is realistic to expect. This does not just underserve ideals of responsible SDM, but possibly undermines them. Persons come to doctors in vulnerable states; they need careful support to responsibly come to what are inherently ambivalent medical decisions. This process entails that doctors and patients learn *from each other*. Instead, law treats patients as autonomous health consumers who need to be given information to support their ‘free choice’ making. In addition, the law does not oblige to explain to patients (does not ‘list’) what the information obligations are for, and therewith gives them little clue about how the testimonial process can support them in SDM. Cited at various points already was how ‘explanation’ is not used at all.

There was a lot of literature for law to go on if they would have wanted to engage with the social dimension of medical knowledge practice better, including literature on how it is a good idea to do so if non-oppressive medical practices are what is aimed for. The domain study cited scholarly and practitioners’ arguments on how after centuries of doing the exact opposite of what SDM aims for, both doctors and patients need capability training.¹⁰⁴² More guidance on how to make knowledge together, and decisions based on that knowledge. Rather than adding more ‘bits of information’, types of discussions could be listed: exchanges that do more justice to the kind of process that (shared) medical decision making is—and, a Model’s view would add, do justice to truths and myths about medical knowledge itself. Compare how medical knowledge (/the algorithm) gets to be referred to as the medical-technical, and as belonging to the doctors’ (/mathematician’s) side of SDM. ‘Values and preferences’ (/social and ethical concerns) in such a presentation are what patients (/social researchers and affected parties) bring to the table. This ignores how medical findings are not self-evident, and need interpretation *in light of* what patients know, think, want, and need. And vice versa: it will be hard for patients to express preferences and values without understanding themselves from a medical point of view.

Suggestions in literature for a move towards a legal explanation paradigm that is less focused on bits of information could be of help here, such as the addition of more ‘care’ related subjects. For example, expanding the obligation to inform about a possible treatment choice to include discussing with patients what it would mean to *abstain* from recommended treatment. Other cited suggestions included obligations to refer to further knowledge about their (predicted) states, and how to handle self-care and informal care relationships.

¹⁰⁴²E.g., there is a backlog of hard to bridge knowledge and understanding gaps; inevitable social power asymmetries potentially make patients into bad self-representatives; there is unclarity about the information patients need to have brought in to establish that the decision was in fact ‘shared.’

The same care should be taken in deciding about what to list for, since it needs to be understood how an ‘item’ will influence practice. An example: the modern trend to add diagnostic results directly to remotely accessible patient files leads to discussions. Learning of (suspected) diagnoses already has mental, social, and even physiological effects on patients, and in addition, unmediated results can be misunderstood. But simply cutting off patient access to such information, as a return to more paternalistic practices, understandably finds resistance too.¹⁰⁴³

6.4.3.2 *The role of law in ‘realizing the best possible care’: arguments for a progressive uptake of professional norms in AI-informed times*

The preceding argument is also an argument for the value of explaining *per se*: for not tying every ‘item’ to a narrow goal but to aim to serve the larger goals of responsible, mutual understanding. This value is much better acknowledged in professional and ethical norms in the field than it is in law. Responsible patient understanding and continuous dialogue was described as instrumental to the establishment, maintenance, and perpetual development of a *responsible* trust relationship.¹⁰⁴⁴ And it matters what kinds of epistemic input and what knowledge co-creation methods are used by doctors towards that aim. There was critique on how EBM’s influence on GP practice (with its quantitative, evidence-based, protocollized care) had pushed aside more patient-centered, and patient-collaborative methods. Arguments for a better balance were made, e.g., engaging patients in the creation of EBM treatment recommendation instruments. This is an important discussion to learn from in a time where data science and AI are delivering ever more quantified knowledge to the mix. As in other decision-making spheres, the *kinds* of claims that such methods can make may come to attract much attention, to the detriment of other kinds of claims that deserve weight in SDM practice, too.¹⁰⁴⁵

For example, research on the role of explainability for fostering trust in medical AI does not necessarily start from a problematized status quo, i.e. from an in-depth discussion of the relation between trust, explanation, and the paradigm of informed consent in relation to SDM and medicine’s problematic histories.¹⁰⁴⁶ Also, questions formulated in such investigations easily start from the premise that AI can deliver unmatched accuracy, efficiency, and personalization compared to a status quo without

1043Section 5.4.5

1044Still ‘awkwardness’ was found in professional norms as well, at least from an outsider’s perspective, such as in professional guidance on the implementation of the (first edition) WGBO: the guidelines for example warn doctors not to assert undue pressure on their patients’ decision making, but also to engage “their full capacities of conviction” if that would be necessary to push a choice that aligns best with a patients’ interests.

1045Green and Chen, ‘Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts’.

1046See for example Julia Amann et al, ‘Explainability for artificial intelligence in healthcare: a multidisciplinary perspective’, *BMC Medical Informatics and Decision Making* 20, nr. 1 (30 November 2020): 310.

this: a framing that Moss et al. discussed as problematic because it negatively taints important critique as a loss of opportunity.¹⁰⁴⁷ Preliminary findings from a study based on in-depth interviews with diverse interested parties (clinicians, medical technologists, screening program managers, consumer health representatives, regulators and developers) suggests that views on the value of explanation for (specifically) diagnosis gradually fall on two sides, *regardless* of whether AI is involved.¹⁰⁴⁸ Interviewees on both sides agreed about the need for human oversight, critical thinking “among clinicians,” and patient safety, but not on the instrumentality of explanation to meet these ends. The authors write how “a different epistemic basis” for trustworthiness seems to underlie each approach, where the ‘outcome-assured approach’ appealed to “evidence and assurance from experts” rather than on a “culture of contestation” across different levels of expertise for the establishment of trust.¹⁰⁴⁹

Those are quite different bases indeed. A Model view would argue that the research underlines how it is important to emphasize how doctors are already (legally, ethically and professionally) obliged, and also already challenged, to help their patients decide ‘whether to invest their trust or not’.¹⁰⁵⁰ To avoid that research outcomes such as these are used to inform the (further development of) professional norms that law now relies on by *toning down* the explanation values that are currently embedded in them, law should take a more pro-active stance, and codify those norms that are already in place, better. For example, law could look at ‘moral consultation’ methods that help GPs discuss particularly value-laden medical situations with each other, and—relevant for this element—with their patients. Such a setting situates the doctor-patient relationship in the larger social-epistemic world it exists in. The method as it was cited entailed a broadening of the deliberative space with other knowers such as (in)formal carers, and other ‘knowledges’ such as norms from e.g. law, ethics & professional standards, and health policy. A comparison with the type of multi-disciplinary, ‘impact assessment’ processes cited earlier could inform to enhance this space with critical AI scholars to help all parties to more meaningfully informed SDM processes.

6.4.3.3 *Honesty in the form of sharing conscientious concerns*

The above-described broadening of the deliberative space necessarily serves the information positions of ‘expert’ decision makers, themselves. Their critical appraisal of new technologies, of decision-making aids/methods, and their further development of professional norms need to be related to explanation obligations to make sure that

¹⁰⁴⁷Moss et al, ‘Assembling Accountability’.

¹⁰⁴⁸Yves Saint James Aquino and Stacy M. Carter, ‘Explanation versus Outcome: Examining Professional Perspectives on the Ethics of Explainable AI in Clinical Diagnosis’, 2022, <https://juanmduran.net/explanation-versus-outcome-examining-professional-perspectives-on-the-ethics-of-explainable-ai-in-clinical-diagnosis/>.

¹⁰⁴⁹Aquino and Carter.

¹⁰⁵⁰O’Neill, ‘Accountability, Trust and Informed Consent in Medical Practice and Research’ See the discussion in section .

patients', and other AI subjects' social-epistemic needs are made to count. To illustrate, the domain study discussed a tension among members of IC-ward teams that were being trained to apply a black-scenario triage decision system during the height of the COVID-19 Pandemic. Some members asked to remain anonymous for the triaged patients since they were uncomfortable to face them. This would illegally leave a life-or-death decision unaccounted for, a severe infringement of patients' right to know who makes decisions about them, and why.

This thesis would additionally argue that from this, it follows that they have a right to a decision maker who refuses to act if they fundamentally disagree with the justness of what they are expected to do, and explains to their patient why this is so instead of 'going stealth.' This important point is also made by Blythe and Curlin. They wrote a critical response to an article that advocated a duty for medical practitioners to either perform novel treatments that are adopted by professional standards but that go against their own conscience, or to refer to a 'care provider' who will. The authors do acknowledge the value of keeping doctors to their norms, and reveal how 'liberating' scientific inventions were obstructed by an unwilling profession in the past (e.g., the contraceptive pill). But they warn for the opposite, which they describe as an all too Weberian, bureaucratic approach. Obliging physicians to perform treatment that is not contra ethical and professional norms (or refer to a doctor who will) disregards how those norms themselves have sustained the worst kinds of scientific "progress." The forced sterilizations and genetic screenings they refer to were only 'unethical ends' in retrospect. "[Individual] conscientious refusals," they write, "alert the profession as a whole to regions of practice that require further deliberation." But, this thesis argues, for patients to benefit, such conscientiousness needs to be upfront and discussed with them, too.

Currently, patients are presenting all kinds of new consumer-health technologies to their doctors, who in absence (yet) of trustworthy standards are shy to act.¹⁰⁵¹ Through medical tech's media strategies, patients are exposed to promises of the affordances of AI and big-data driven, personalized health care, for example in the form of commercial diagnostics. The push on physicians is therewith large from all sides, and a 'tech knows best' paradigm looms. Medical decision makers could improve patients social-epistemic information positions by discussing warnings from the past & present with them: how 'know it all' GPs missed diagnosis because they assumed too much knowledge where in fact they lacked crucial, qualitative input from patients. Input that is not substituted by gathering more data and information about them, whereas 'more data' is precisely the fix that the AI field likes to propose in order to deal with fairness

¹⁰⁵¹Silven et al, 'Clarifying Responsibility'.

issues (racism, discrimination, intersectional effects).¹⁰⁵² The earlier cited argument of Moss et al. who warn of the limited imagination of impact assessments that compare risks & benefits of a plan to a counterfactual world where the projected benefits are not realized¹⁰⁵³ is relevant here too. AI impact assessments could deliver useful information for explanation practices, but especially if they help to understand how a new affordance is part of a continuum of ‘new’ medical developments.¹⁰⁵⁴

6.5 Creating records (element four)

6.5.1 General observation

The fourth element obliges to create records of the testimonial exchanges that allow to assess how all modeled duties were attended to. The idea is that these records do more justice to how explanation is a knowledge practice, itself; one that itself can therefore harm or further participants’ ‘meaningful information positions.’ Such records, aggregated, also support studies of explanation practices.

The thesis did not expect to find many legal rules that oblige to create records of explanation practices in quite the ‘meaningful’ way that element four requires, but even lower expectations were disappointed. This is concerning in light of how AI-infused decision support systems are trained on evidence of past decisional outcomes.

6.5.2 Observations from the Administrative domain

Of the type of decisions that were focused on, most initial decisions are sent by post, accompanied by a minimal set of reasons. They are not discussed at all with explainees. By asking for an account of the testimonial exchange about the outcome of a decision process, element four therewith asks for something that is currently not part of explanation rules in the domain. That does not mean that no useful ‘testimonial’ information is put to record; in compliance with due process and archival duties such information will be added to the Administrative case file. But these are not supplied to subjects as a rule. Different purposes drive the selection of that information, and not all records of interest are an obligatory part of them. For example, in the phantom vehicle cases, written correspondence about (the outcome of) a process was kept, but not the phone records of victims when they submitted additional information. This is unfortunate also in light of how ‘phoning in’ remains especially important for the

¹⁰⁵²As one AI creativity researcher stated, “You don’t need to build a bigger nuclear bomb to know we need disarmament and missile defense. You build a bigger nuclear bomb if you want to be the person who owns the biggest nuclear bomb.” Alex Hern, ‘TechScape: This Cutting Edge AI Creates Art on Demand – Why Is It so Contentious?’, *The Guardian*, 4 May 2022, sec. Technology, <https://www.theguardian.com/technology/2022/may/04/techscape-openai-dall-e-2>.

¹⁰⁵³Moss et al, ‘Assembling Accountability’, 25.

¹⁰⁵⁴See also section 6.3.3.1 ‘times are different now?’

digitally challenged, as it already was for the ‘bureaucratically challenged’, a very large group.¹⁰⁵⁵ Their expressions of desperation, their need for help to correct their information, and their lack of understanding of decisions remain outside of the case files, therewith risk to be ignored in (internal/judicial) assessment and aren’t available for research. Of equally problematic status is the fact that the different reasons (and even decisions) that administrative bodies produce in review procedures are not necessarily recorded. E.g., when an unchanged decision is grounded on a different set of reasons, this may not be known on the basis of an individual case file – let alone from (aggregates of) public records.

Accounts of how decisional outcomes were reached and discussed are therewith incomplete. They are of insufficient use for explainees who need be able to engage with (formal and informal) others about how and why they were treated this way. The accounts are also of insufficient use for the necessary assessment and improvement of decision and explanation practices: to serve progressive understanding. The gap can only be filled by doing empirical, including ethnographic work. This is taxing, and the outcomes cannot be related to goals that are not currently codified. Nonetheless, studies that were done in the past have invariably argued that decisions and explanation practices are both in need of improvement even to comply with contemporary standards.

The thesis has argued that the development of legal standards for ADM explanation should not start from suboptimal baselines in ‘good old’ explanation rules, nor assume that the current state-of-art in explanation practice is a benchmark to strive for. Decision support systems are already being trained for use in legal and other environments on what is made available in machine readable ways about the process and outcome of past decisional processes; including records of how the outcomes are reasoned. To illustrate, Saxena et al. (cited earlier) studied a system that was being trained on the output of a quantitative instrument that was used by case-workers to determine a family’s risk score. But the case files also included hand-written case notes, based among other things on colloquial exchanges with families. These too could have been made machine readable. It just takes more time, and includes a qualitative coding step—it requires to include more kinds of expertise in the system design team. The authors did this, and revealed salient human and bureaucratic influence on the outcome that could not be gleaned from what was studied before. Among these were individual and institutional instances of oppressive knowledge and decision making.¹⁰⁵⁶ One important finding for this thesis’s purposes were power dynamics that clearly subverted the pro-equality, ‘families as partners’ goal of the particular administrative body. Such findings are not unlike concerns about the legal support for SDM that failed to take the social power relationship of doctors and patients into account.

¹⁰⁵⁵ ‘Weten is nog geen doen. Een realistisch perspectief op redzaamheid’.

¹⁰⁵⁶ And useful clues for what would need to be further researched to understand more about the actual decisional process, such as dimensions of human-machine interaction. Saxena et al, ‘How to Train a (Bad) Algorithmic Caseworker’.

Relatedly, in their annual report about 2021, the Dutch National Ombudsman argues that aggregates of (in their case) individual complaints procedures about administrative behavior are an important source for administrative bodies to learn *about* their institutional practices. These reports played a salient role in uncovering practices of what later became acknowledged as types of ethnic profiling that should be avoided even if the Court had refused to condemn the practice.¹⁰⁵⁷

Another illustration pertained the development of an automated decision support tool to assist Administrative Court's paralegals when preparing WAHV case files for court procedures, including suggested verdicts: these are the Traffic Law cases that the phantom vehicles cases are part of.¹⁰⁵⁸ Per the report, the “most important factor” that the paralegals take into account when preparing cases for court hearings is the appellant's motivation. These reside in different places, and come in different formats for organizational, digital and document related reasons. Handwritten letters are among them, the use of language by appellants and/or their legal representatives is highly diverse. Sometimes no motivation is found. Since for these reasons no “common pattern” could be defined, the system was built to extract “the essence” of the appellants motivation as represented in the public prosecutor's reasons for denying the appeal.¹⁰⁵⁹ Since this is a choice that harms the rights of defendants, a refusal to automate that part of the case file, a refusal to impoverish that information feed, would have arguably been the more responsible choice.

6.5.3 Observations from the GP domain

The studied explanation rules (legal and otherwise) of the GP domain contained no obligations for the creation, and provision, of explanation records. Per the WGBO's rules, ‘written or electronic’ information that was discussed is provided upon patients' request only. Whether patients consented (and so, were informed) legally needs to be recorded in their patient files, and some records of what information was provided will be in there if only to sustain complaint, fault, and malpractice proceedings. File-related duties were however not studied – it would be advisable to do so for a more comprehensive understanding.

It was however clear that several kinds of things that could usefully sustain such proceedings are not necessarily recorded. These are things that would also be valuable material for those who are currently engaged in AI explainability design for Health Care. Records that demonstrate how doctors do, and do not, conduct ‘interactional justice’ in terms of the model will add value to what is already available through

1057 ‘De burger kan niet wachten: Jaarverslag van de Nationale ombudsman, de Kinderombudsman en de Veteranenombudsman over 2021’, 21.

1058 Narayan, Nitin ‘A Decision Support System for the Court of East Brabant’ (Professional Doctorate in Engineering, Den Bosch, Jheronimus Academy of Data Science, 2019).

1059 ‘A Decision Support System for the Court of East Brabant’, 30, 38.

empirical research, patient communication studies, et cetera. One such thing is a record of whether patients understood what they were informed about: the critiqued omission in the WGBO's update.

Another suggestion follows from the context of electronic patient file developments that were cited earlier. Doctors are concerned about the cognitive and emotional effects on patients who are confronted with unmediated medical findings. At the same time, a particular recommendation of the Medical Association to sustain patients in their dealing with *mediated* consultations by encouraging them to make recordings is met with mixed feelings by medical professionals. The recommendation is informed by the emotional and cognitive patient understanding challenges during consultations, including memory challenges, but could also serve Element four's purposes when such were added to the electronic file. This too is met with less than enthusiasm by a large percentage of doctors however.¹⁰⁶⁰

Another type of record that would be of use and that is indeed being created after a change in governance, are the outcomes of GP (and other carers) complaints procedures. Interesting things surface from these records. For example, the current aim of GP procedures is 'reconciliation' rather than resolution. But reconciliation may not be the right goal to strive for when it is not yet clear whether a complaint is about the content of care, or 'just' the communication of it. It is acknowledged that this is frequently hard to distinguish.

6.6 Mobilizing observations for AI-infused times

This last part of the chapter assembles the essences of the technology-related observations that were made in the preceding sections, and adds some thoughts to these. Categorized succinctly under each Model element, the exercise provides a short and accessible, yet incomplete oversight of how the thesis's lessons are usable for explanation (re-)design in AI-infused contexts. Instead of the shortened Model descriptions that were used in the chapter, the original descriptions will be used to close the thesis. Element one starts with a very brief recap of why there needs to be this kind of attention for explanation in AI-infused times. And again, the elements build on each other: neither of them 'work' without the other.

As stated earlier, the Model is meant to be further developed for in-context use through multidisciplinary investigation. Some clues on how to do that are included below. So what is the point of stating, seemingly superfluously, that the advisory statements are incomplete? For one, it is a nod to the author's reluctance to 'finalize' accumulated insights, e.g. by modeling them into digestible chunks. This can certainly be seen as a flaw, since it leaves readers with more work to do in applying the work that was

¹⁰⁶⁰Section 5.3.4.2

done, and less guidance to do it. In a thesis that among other things advocates to share explanations in usable ways, that is ironic. Hopefully, this section takes away some of that pain. Still there is a point to this refusal to make final points. The introductory section on the methodological approach (section 1.2.2.3) voiced the thesis's ambition to provide something more than 'impressionistic guidelines', but—cf. Lorraine Code—still less than 'necessary and sufficient' guidance. The first aim was to help readers imagine what an explanation practice that aims for knowledge justice could look like. The second, having modeled the appropriate values, to categorize Model-based findings in a way that helps this process along in AI-infused times specifically. Let's call it 'concrete guidance.'

6.6.1 One

First duty, or element one: investigating explainer authority

Explainers are obliged to investigate their own social-epistemic positions with regard to their decision-making modalities, and their domain's underlying (input) knowledges in order to assess their role authority: does the explainers' understanding justify their authoritative and trustworthy explainer position? If no (or can't investigate), rebel.

This element obliges that explainers avoid to become an instrument of unjust ('bad', oppressive) knowledge practices, and are able to explain their 'avoidance strategies' to their explainees. To what extent they need to in fact explain these strategies is best determined in a decision domain's context. More positively expressed, this element promotes that explainers are able to communicate how, and not just that they are trustworthy 'knowledge practitioners,' and not just accountable decision makers. The point at this stage is to link the self-reflection of explainers to their position of authority vis-à-vis explainees, as part of responsible practice. The need for explainers to rebel exists when explainers feel incapable to do this, for example because they don't have access to justificatory sources or aren't afforded the time, means, or authority to investigate.

Addition: those accountable for a particular decision context are obliged to ensure that explainers are epistemically equipped to recognize oppressive dimensions of a context.

*

The oppressive potential of fast proliferating ADS across decision contexts, and the weakness of legal protections in place make the risk to become an instrument of bad knowledge practices arguably large in AI-infused times. As was argued, AI's consequential problems of interrelated racist, discriminatory, and marginalizing nature are a bad fit for a human rights framework that has failed to apply itself to "the logic

of advantage/disadvantage”¹⁰⁶¹ as it manifests in our societies. This also expresses in how problem descriptions of AI’s challenges, not least for the right to explanation have been framed and responded to. More generally and perhaps most fundamentally, the understanding of fairness, and of algorithmic fairness in its wake tends to be aligned with a legal anti-discrimination framework that already underserves many people.¹⁰⁶² But as Moyn writes, this poor performance was neither inevitable nor in line with the more ambitious, equity-oriented aspirations in human rights’ history. To improve the situation, the demand of human rights needs to shift: from asking those in and with power to act “more humane,” to the demand for a more just distribution of powers itself.¹⁰⁶³

This thesis has focused on knowledge making powers specifically. AI is of increasing influence on facts, norms, and concepts used in laws, medical and other expert knowledge, and all kinds of policies. Yet the governance orientation with regard to ‘algorithmic’ transparency, accountability, and explainability has been on decision making rather than (the) knowledge making (that goes into this), and has done this built on a poor understanding of the knowledge making characteristics of explanation processes themselves in addition. Element one would advocate to take advantage of an identified momentum to ‘up the game’ for tech-neutral as well as tech-oriented explanation regulation, and to use the modeled insights to do this conscientiously. Several examples follow below.

In the Netherlands this momentum is found for example in the aftermath of (bureaucratic, political, judicial) justification and explanation failures of the Benefits Scandal, and in the Senate’s acknowledgements of the failures of anti-discrimination legislation to respond to with institutional racism and discrimination that is embedded in legal assumptions and their policy translations. However, the chapter also found that the Senate’s expectations for legal intervention into oppressive knowledge practices are set rather low. Their report considers how law is unlikely to change causal attitudes, morals, and culture; it paints civil servants as non-obvious perpetrators with other priorities such as dealing with Government’s “mixed messages” (i.e. to keep everyone to the same rules but do individual justice – the so-called Weberian dilemma). The chapter considered how their advice to instruct administrative decision makers to engage more strongly with “the spirit of the law” disregards law’s spiritual problems. This adds to the meagre State acknowledgement of the existence and influence of institutional racism and discrimination. It was late, legally misguided, and therewith

1061 As Hoffman writes, “we need to broaden our scope to better account for the (re)production of the full range of social hierarchy – that is, we must move beyond analyses that center and scrutinize conditions of relative disadvantage to also account for the normalization and production of systematic advantage” Hoffman, ‘Where fairness fails’; Williams et al, ‘Surfacing Systemic (In)Justices: A Community View’.

1062 Hoffman, ‘Where fairness fails’ The same can be said for the alignment with existing ethical principles for fair decision making that have not explicitly embraced a more relational understanding of human flourishing or (at least) distanced itself from understandings that sustained the opposite.

1063 Moyn, *Not Enough*, 217.

misleading for decision subjects. These observations are important in light of the easy referral in ADM explanation strategies of ‘meaningful’ oversight moments for humans in the loop of decision practices.

More generally, momentum was found in critical AI literature and how this is finding its way to the decisional contexts the thesis is concerned with. But this happens without a negotiated strategy for which insights end up being used and interpreted for which knowledge bases (legal, ethical, professional normative frameworks) in use by State and other decision makers, where they come to inform designated explainers. An anecdote that seems of little importance at face value is used by way of illustration. Among its tailor-made resources, the Municipality of Amsterdam counts a *Fairness Handbook* for the design of in-house AI systems, “an A-Z manual to measure how fair your model is and to mitigate the biases you encounter.”¹⁰⁶⁴ A quick scan of the 60-odd pages booklet reveals a broad but somewhat haphazard collection of relevant research for those who want to avoid that their AI becomes part of oppressive knowledge (making) systems. The way the insights are applied, gathered, modeled and interpreted arguably falls short on various salient points.¹⁰⁶⁵ With that, the status of the booklet as a norm setting document becomes more important. In a meeting with a working group for municipalities around the establishment of a National Algorithmic Register,¹⁰⁶⁶ the city’s *Fairness Handbook* was cited as a fond resource for developer teams and civil servants. Yet neither the authoring organization’s website nor (when asked) municipality explained the booklet’s status.

Momentum was also found in various approaches to human/fundamental impact or risk assessments for ADS. These are not yet broadly obligatory at this time,¹⁰⁶⁷ they are *promoted*, and come described as useful ‘discussion tools.’ One question is to what depth these discussions are meant to be taken. The Fundamental Rights and Algorithms Impact Assessment Tool (IAMA) that was adopted by the State writes how IAMA assessments can function as an overarching tool in which other assessment tools,

1064 Selma Muhammad, ‘The Fairness Handbook’, 17 May 2022, <http://amsterdamintelligence.com/resources/the-fairness-handbook>.

1065 Saliently, the booklet conflates fairness with the notion of equal treatment in the legal antidiscrimination framework; it overestimates the affordances of ‘de-biasing’ methods, and presents categorizations that are limiting (e.g. only identifying nationality and postal code as origins and proxies for discriminatory school admissions, ignoring how other example variables, i.e. grade point average and extracurricular activities are proxies, too) or confusing (e.g. between fair allocation of services, and other qualities of services). Last but not least, explanatory capacities of the AI are only introduced in “Phase 5: Implementation & Deployment” even though the booklet states how a lack of explainability make (the earlier planned) fairness assessment impossible.

1066 Soon to be made obligatory, for example based on the EU’s proposed AI Act See e.g., <https://vng.nl/nieuws/gemeenten-starten-met-een-algoritme-en-sensorenregister>. The beta version can be found here: <https://www.algoritmeregister.nl/algoritmes> Meetings were attended by the author on October 31, November 14 & 28.

1067 Although for the more impactful algorithms this is to be expected, see e.g., a recent Parliament initiative. ‘Motie van de leden Bouchallikh en Dekker-Abudlaziz, Kamerstukken Tweede Kamer, vergaderjaar 2021–2022, 26 643, nr. 835’, March 2022.

including those part of a law's or a policy's preceding 'analog' democratic legislative process but also other tech-oriented tools are explicitly embedded.

This is where the Senate's remark that law should not be expected to change causal attitudes, morals, and culture is a disturbing factor. Wrongful attitudes, morals, and culture are embedded in the laws that the IAMA-assessed algorithms are based on, and the implicit signaling to decision makers/explainers (the 'non-obvious perpetrators') that technology can safely be applied while we patiently wait for laws' moral updates and for organizations' cultural awakening is irresponsible. An illustration: one civil servant working with the IAMA to make the city of Rotterdam's fraud detection algorithms safer acknowledged the harshness of the welfare laws they are built to implement. Yet they argued that algorithms "only give directions" for case workers to follow up on for fact finding missions.¹⁰⁶⁸ This view ignores e.g. how wrongfulness establishes in the type of knowledge making that is used and the directive role it is given, and fails to acknowledge how giving a certain role to *this* kind of knowledge in decision making can come at the cost of initiating other kinds of knowledge making for policy, similar to how the quantified knowledge of EBM was seen to take up space at the cost of qualitative methods in Health Care. The example sustains the addition to element one that was suggested in section 6.2.4. The concern addressed in that section was the possible lack of explainer capability to recognize knowledge related oppression, either because of explainers' own social-epistemic embeddedness, or lack of understanding of technological histories and methods of oppression.

More generally, for element one, the example is significant in light of the big role that algorithmic assessment processes are expected to get. E.g., the burden of checking procedural rights compliance with regard to what happens during decision making is shifting to precautionary tools before (such as assessment procedures) and ex-post judicial oversight. There, effects rather than procedure are expected to become the focus of scrutiny, taking up space at the cost of causality.¹⁰⁶⁹ The first instance explainer in the middle risks to be left without the necessary capabilities. As was argued, useful clues for responsible explanation can certainly be produced through assessment processes, but not when the roots of wrongfulness, alive in 'the spirit of the law' and the spirit of other bodies of 'input knowledge' are to be considered out of scope and digital methods are not sufficiently understood. (More requirements are pointed out under later elements.)

1068'Interview: "Laten we nou vooral leren van gemaakte fouten en kijken of we algoritmes wél verantwoord kunnen inzetten", *Verdieping - College voor de Rechten van de Mens* (blog) (Ministerie van Algemene Zaken, 7 July 2022), <https://www.mensenrechten.nl/actueel/toegelicht/interviews/2022/laten-we-nou-vooral-leren-van-gemaakte-fouten-en-kijken-of-we-algoritmes-wel-verantwoord-kunnen-inzetten>.

1069Binns and Veale, 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR', 332.

When assessment processes are designed as highly reflective knowledge making practices rather than just as assessments of decision practices, their potential to progress a domain's explanation practice is more realistic.¹⁰⁷⁰ They could infuse domains with the kind of knowledge making that it requires, including training for the kind of explanatory conversations that have not been happening in them. It is wise to reiterate here how it is advised to engage with the problematic history of the 'technological complexity argument' and its AI-inspired revamp as was discussed in this chapter. Also, possible knowledge deficits of a domain's explainers with regard to the types of epistemic wrongs, harms, and dynamics to look out for need to be investigated and engaged with.

A second question is what authoritative conclusions are to be drawn on the basis of the outcome of assessment processes. Their status is yet to be firmly established. Laws are made through democratic processes; other bodies of knowledge have their own established methods. What status can be given to the knowledge that comes out of, say, a more progressive human rights assessment? What should that status depend on, and whose decisions are this? The GP domain study showed that law has tended to shy away from adopting meaningful ethical and professional norm setting. Framed mainly as not wanting to intrude upon medical professional circles, there is also evidence of the reverse: the Administration domain clearly showed a reluctance to adopt progressive insights from scholarly, ombudsman, and other sources; and (Dutch) politics responded fiercely to the multidisciplinary COVID-19 IC bed scarcity triage model that originated from the medical domain. When human rights assessment procedures are given stronger legal status and the questions raised above are meaningfully resolved, explainers in AI-infused times stand to have a stronger position from which to be critical of higher authority, but this is not yet the case.

6.6.2 Two

Second duty, or element two: engaging with the social-epistemic positions of explainees

Explainers are obliged to investigate the social-epistemic positions of explainees in relation to the decision-making modalities and underlying (input) knowledge at hand; can explainees be expected to responsibly provide (or have provided) the necessary input, and understand the output? If no (or can't investigate), rebel.

This element, like element one, obliges to 'prepare the table' for the negotiation of the how's and why's of decisional outcomes. This time the focus is on how explainees will be able to experience a just testimonial process. Explainers need to be able to

¹⁰⁷⁰Such as the Ada Lovelace example, which was characterized above as to "promote to expose typically (and problematically) separated parties to each other's expertise, to redistribute social-epistemic authority, to infuse the process with critical wisdom from outside of the development circles, and to support the process with multidisciplinary guidance." 'Algorithmic Impact Assessment'.

demonstrate engagement with their explainees social-epistemic situatedness (on individual and group levels) with regard to the larger decisional process and methods: 'the system.' This includes engagement with how a system historically treated explainees as a group and individually. The need to rebel exists when explainers feel their explainees are in no position to participate in the decisional process responsibly.

*

For element two, the failure of the Human Rights framework and its codifications to understand and protect humans in terms of their relations, and in terms of their entanglements in institutional power structures is exacerbated in AI infused times. The sheer amount of (public, private, entangled) data relations, and the way these relations establish cannot be meaningfully understood by explainees and their explainers. The EU's boasted human-centric data and AI regulation frameworks have not stood in the way of these developments (and have arguably added their own bulk of information that is increasingly challenging to keep track of.) By aligning the extent of automation and its technological complexity to the extent of potential harm, the frameworks also struggle with the blurred lines between manual and automated steps in decision making. Last but not least, the digitalization of decision environments (and society in general) has proven challenging for large groups of people, worsening their capabilities for responsible participation. In terms of engaging with explainees, better, a conclusion that follows from all this is that there is a lot to be won but times have become harder for winning. It is therefore important to reiterate some 'lessons' from the preceding chapter here and build on them some more.

One thing that needs repeating is that simply creating more discretionary space for decision makers does not naturally lead them to engage with citizens in ways that allow for better informed participation in decision making. It was even established that the opposite happens: obscure policies are created and even litigated for in court by administrative bodies. And, as always, a 'doctor knows best' scenario looms when dependence on human explainers replaces responsible explainee understanding.

A warning also pertains to focusing on review procedures as a way to protect persons' rights when harms that they experience are not adequately captured by 'algorithmic governance.' Review procedures should not be used to identify harms that can be identified in earlier stages (although that is obviously a quality of them). This unfairly burdens explainees; *un*burdens initial decision makers/explainers which is contra what these times need; and such procedures rely on existing legal definitions of harm which are too individual of character.

Assessment procedures, again, can help to improve. What needs to be part of such procedures is to take the decisional domain's cultural and historical context and justice failures into account; to investigate the impact of the individual decisions a system will

(help to) make in terms of the larger institutional context that applies to prospected explainees; and to study the interplay of qualitative, quantitative, and automated methods at use in the assessed decisional process.

The engagement *with* explainees themselves in these processes is crucial for identifying their responsible participation affordances. E.g., since conversations where responsible explainee distrust is on the table have not yet been promoted by legal explanation rules at all, such engagement can prove to be meaningful training ground for explainers and explainees both. However, a system's (projected) affected groups of persons should not solely be burdened with establishing evidence for the beneficence and maleficence of a proposed ADS. The point first needs to be to gauge prospective explainees' grasp of the larger system. They may have more insight from experience than a domain's explainers have, but also less knowledge of what happens on levels that are not accessible to them. The Benefits Scandal, and much cited literature (notably, Eubanks) showed how harm ensues when either (or worse, both) are ignored by decision makers.

The process should also avoid to exhaust their resources and (financially, socially and practically) facilitate their participation in the processes. The cited advice(s) to let explainees bring buddies to Administrative and Medical decision and explanation processes can be adopted here, too, and would prevent that a false sheen of horizontal relations (since, there is no hierarchical decision making going on *yet*) negatively influences the type of knowledge making that can happen in them. Last but not least, affected explainees will (first) need to be found. This takes an effort that goes (way) beyond, for example, publishing planned assessment procedures in Algorithmic Registers under a header called 'civil participation' or a feedback form. Connection needs to be sought with publics and individuals who are not yet aware of the impact a (type) of system may have on them and with groups that face barriers for getting involved.¹⁰⁷¹

That said, national (and local) registers can certainly come to function in a way that explainees' information positions are *generally* improved, and the same is true for explainers: the Municipality of Amsterdam mentions how the second main targeted group of users are civil servants. It means to guide them with regard to "what kind of transparency is needed and how to provide this information understandably."¹⁰⁷² A lot of decisions are yet to be made in and about them, and will influence the value of the registers for both (and other) parties. One set of decisions is related to the question which algorithms to publish. High-to-low risk triage systems are already being designed, and entail decisions on what impactful or 'risky' decision making

¹⁰⁷¹ Moss et al 'Assembling Accountability'.

¹⁰⁷² Meeri Haatja, Linda van de Fliert, and Pasi Rautio, 'Public AI Registers: Realising AI transparency and civic participation in government use of AI' Whitepaper written by Meeri Haatja, Linda van de Fliert and Pasi Rautio' (Gemeente Amsterdam), last consulted 30 November 2022, <https://algoritmeregister.amsterdam.nl/wp-content/uploads/White-Paper.pdf>.

is. Warnings from the Administrative domain apply: don't simply relate the lack of complex digital methods (i.e. 'simple automation') to low risk decision making.

The bulk of decision making is about what to publish about the selected algorithms, though. It would arguably help if the algorithm registers set an example of what it means to present oneself as trustworthy explainer, not just in acknowledgement of past algorithmic mishaps but in acknowledgement that large groups of citizens have older reasons to mistrust the State, their Health Care systems, and other decision contexts. Some considerations towards this are listed here.

1. For public decision making, the 'legal basis for decision making', i.e. law and policy are likely to be listed. Earlier concerns about obscure policy apply, and (so) it would help to elaborate how an algorithmic method is seen to best support a particular policy goal, how that policy goal is sustained by law, what legal protections apply in case of harm and how these are indeed adequate. Referrals to e.g., 'principles of public administration' or legal anti-discrimination frameworks do not meet this mark. 2. It would also help to elaborate on cancelled systems: why and at what stage of (either) planning, design, development, testing, or implementation were they aborted? A 'kind of' Parliamentary history, if one will. 3. Novel instruments that are referred to, such as risk assessment procedures, certification schemes, need to be explained. 4. A question raised in the working group on the National Algorithmic Register pertained to 'underlying algorithms', such as those used by municipalities' information headquarters and the 'care fraud information hub.' The question is fundamental and interesting, and argues for a very clear explanation of the 'depth and reach' of the register.

6.6.3 Three

Third duty, or element three: practicing interactional justice

Explainers are obliged to practice interactional justice, which entails to recognize explainees as knowers and rights-holders. Explainees should be provided information that is proportionate to their pre-investigated and incidental (self-expressed) needs; their knowledge and understanding of relevant, larger & smaller knowledge making processes at hand should be discussed with them with the aim of promoting their responsible (dis)trust; accessible justificatory sources from outside of the authoritative setting need to be pointed out accompanied by instructions on how to follow up on such leads; explainees need to be afforded information about their rights with regard to the explanation and the decision outcome; the possibility of social pressure needs to be mitigated by e.g. allowing to bring allies or make recordings.

The duties of this element describe the interactional dimension and behaviors that need to be given an explicit place in the testimonial process. If any description goes beyond what a process is seen to need, this will need to be justified in the testimonial record.

The inclination of lawmakers to treat much practiced (or ‘bulk’) decisional processes as simple, self-evident, ‘routine’ and predictable has led to sub-optimal explanation practices. The implementation of automation in such cases exacerbates the problems while obscuring their origins.

*

In both domains, law’s weak promotion of what element three obliges to do, despite abundant evidence that the modeled values need strong promotion, stood out. Again, this means that affordances of contemporary explanation processes are insufficient at a time when explainers are more in need of guidance than ever, and in a time where explainees are in (even) weaker power and information positions than before.

The discussions under the previous elements already highlighted various paths towards more meaningful explanation processes and warnings about things that would prevent them from establishing. The sections pointed forward to some important things that are listed here. Among them were the need to engage with pre-investigated needs, considerations about referrals to justificatory sources, the cultivation of responsible dis/trust, and adding companions to explainees to mitigate power imbalances. The advocacy here therefore mostly ‘rests its case.’

One obligation of element three that needs stressing still is to provide explainees with information *about* their explanation rights: what these mean to achieve for them, and what defines success. A meta-obligation, so to speak. And a very important one in a time where there is hardly anything about explanation that is not being re-discussed, re-defined, re-weighted and re-appreciated.

The thesis has welcomed the promotion of what were cited as ‘knowledge making and awareness raising’ sessions that are argued for in algorithm assessment literature, but advises strongly that ‘explanatory clues’ that are raised in them are related to an assessment of explanation rules in place for a decision domain that the assessed algorithm will function in. The gap between what the rules demand to provide and what is and can be made explainable can be expected to be very large, and several tendencies were identified that give no reason to expect that explanation procedures stand to be expanded, even with the GDPR in place. ADS are for example looked to with expectations of increased efficiency—of less need to take more time with (and for) individuals. Cited earlier were tendencies to look to *ex ante* and *ex post* rather than ‘in the middle’ for justification of decisions. Techno-centric arguments about the unusefulness of inter-human and causal understanding apply as well. And, in good old-fashioned explanation regimes, the more elaborate explanation goals are hidden from plain view in legal, ethical and professional principles rather than publicly available instructions. The proliferation of *more* principles was criticized for this reason at several points.

To reiterate here as well is the lack of training for the kind of conversations that the thesis argues need to be had, and how this applies to explainers and explainees both. The health care context history teaches how talking about knowledge making has been as necessary as it was avoided, and for example stressed how ‘medical technical’ findings are not self-evident: they need interpretation in light of what patients know, think, want, and need. Vice versa, it will be hard for patients to express preferences and values without understanding themselves from a medical point of view. Translated to the AI-infused context, people will have a hard time expressing their understanding needs if they don’t know how they are being understood algorithmically. For that, they don’t need mathematical training, they need well-informed honesty. This can only be given if affected publics have been involved in the identification and description of a system’s affordances in the first place, see earlier elements.

6.6.4 Four

Fourth duty, or element four: creating records

Explainers need to create records of explanation practices. These should be understood as truthful accounts of the testimonial exchange as it was prescribed under element three. Therewith the record should express how all previously described duties were attended to, or provide reasons for when they were not. The records need to be shared with explainees, and made available for outside scrutiny in accordance with rules that govern the decisional domain and relevant privacy and data protection regulation.

These record-related duties are meant to produce more comprehensive accounts than the ‘statements of reasons’ that are typically the outcome of decisional processes. This acknowledges how explanation is a knowledge making practice itself, and therewith a place or conduit of possible oppression. Comprehensive records can sustain progressive development of decision and explanation practices across time and domains.

*

Earlier on in this chapter, a ‘general observation’ for element 4 (section 6.5.1) considered how even the low expectations of the thesis with regard to legal record related explanation rules were disappointed. It was mentioned there how this was concerning in light of how AI-infused decision support systems are trained on evidence of past decisional outcomes. This saliently includes records of explanations such as Administrative and Judicial motivations. The same concern theoretically applies to records on e.g., medical informed consent, but specific obligations for medical record keeping were not studied. An interesting aspect for such records is whether it is obligatory to take note of explainees’ wishes *not* to know, *not* to learn of a particular medical aspect that concerns them, or their wish *not* to read a medical expert’s

diagnostic report before it is sent on to another expert. Such information is crucial to understand explanation as a process and as a knowledge making practice.

Without the availability of machine-readable records of explanation processes as a locus of knowledge making, systems are trained to correlate available categories of explainees' personal and contextual data with decision outcomes. This produces even less 'meaningful' information than what could be gleaned when more factors of influence on those outcomes were considered. As just two cited studies already showed,¹⁰⁷³ the inclusion of process-related records will lead to a better understanding of the knowledge that is 'taken for granted' by systems that feed on it, saliently including information that relates to the justness of the process: whether explainees are indeed respected as knowers and rights-holders. That does not take away other concerns about e.g., the predictive quality of 'predictive' ADS, of course.

A second reason for added record duties in AI-infused times is that decision environments have become 'mixed methods' environments. But rather than treating them holistically, explanation regulation is increasingly treating them either separately or refers back to 'manual' explanation rules in place as safeguards of accountability: see for example the notorious reliance on review. The burden of filing for review for explainees was problematized before. To mention here is how they need meaningful records to share their *burdens* of understanding with peers and possibly experts who can support their *responsible* understanding about whether to file for review in the first place. When they do, reviewers need access to sufficiently meaningful records themselves.

This is even more pertinent in times where the amount of people who are unable to 'connect' has grown. The analog literacy of people in the Netherlands already deserves more attention than it gets, and the inscrutability of digital society and/or the lack of digital affordances has added other groups. Such groups need to be able to seek contact live, (with help) in writing, or by phone, and their self-reporting needs to be made to count. In the phantom vehicle cases, this was purposefully avoided for wrongful reasons.¹⁰⁷⁴

What also bears repeating is how 'informal' review methods are promoted with even more expectations in AI-infused times, precisely to provide a more accessible environment. Informal experiments however led to unsafe practices in state-citizen relations.¹⁰⁷⁵ The lack of record related duties for such procedures is a factor that helps to turn informal situations into possible loci of power abuse. The chapter therefore labeled the dichotomization of 'formality' and 'humaneness' as unhelpful. As Moyn writes, "Equality was never achieved by stigmatizing governance but instead by enthusiasm for it, and even devotion to it."¹⁰⁷⁶

1073 Saxena et al, 'How to Train a (Bad) Algorithmic Caseworker'; 'A Decision Support System for the Court of East Brabant'.

1074 Section 4.2.3.6

1075 Section 4.4.3.5

1076 Moyn, *Not Enough*, 219.

Care to explain?

7 The dissertation in a nutshell

7.1 De-idealizing and re-idealizing explanation rules

This thesis conducted research on the legal governance of individual human explanation duties in, and for, AI-infused times. This governance is in motion. That is to say, some of it is. In response to what are seen as game-changing developments in ‘artificially intelligent’ decision systems (ADS), new explanation rules are enacted in law and other instruments of governance. Especially in the European regulatory space, a foundational sensitivity to what is typically referred to as ‘automation’ of decision making is seen to have triggered a righteous legal response to protect individuals against the worst that data processing has to offer: the objectified, oppressive (group) treatment of humans and all the harms that come with such treatment. Especially if it is institutional, especially if it is driven by wrongful ideologies. Modern methods of automation allow to hide such treatment from scrutiny and disable legal protections that are in place against it.

Under the flag of transparency, obligations to explain such ‘black box’ decision making are presented as an antidote. Explainees are given a right to information that needs to be as meaningful as what they already had a right to in established law. The right includes interaction with a knowledgeable human. The idea here is that personal and insightful treatment guards the humane and dignified character of decision processes with weighty consequences for individuals, and therewith the affordance of law as an instrument against oppression.

The narrative expresses a reliance on ideals and affordances of legal explanation regulation ‘as we know it.’ But while regulatory firefighters were (and are) dealing with this tech-induced explanation crisis, a different crisis unfolding in The Netherlands testifies to how the narrative is itself in need of explanation. A constitutional crisis was called after hard won evidence revealed how tens of thousands of families were cheated out of their livelihoods as a consequence of being wrongfully flagged as Childcare Benefits fraudsters. Discriminatory and racist notions in underlying laws, policy and methods drove a boat that sailed on for some 15 years. Some boat hands performed their acts in full awareness (e.g., civil servants deciding to select parents on ethnicity), most of them functioned as instruments of oppression in less overt, less mindful, more cultivated ways. The explanation regimes that governed administrative decision processes had failed to bring any relevant facts to light, and in how judicial scrutiny failed to qualify these outcomes for what they were, Judges’ motivations “reasoned away foundational principles” of proper administration.

It is just one example. Revelations of ideologically tainted (racist, sexist...) medicine are another: they are historically rife, their discovery is hard won, yet the medical ‘informed consent’ is referred to as a foundational explanation paradigm that guards the dignified treatment of patients. With these and other examples, the thesis discussed how decision subjects in less privileged societal positions already suffered the adverse effects that tech-oriented research revealed as alarming developments. Prominent among these are reduced capabilities for responsible understanding and therewith participation, unfair treatment at group level in ways that bypass (awkwardly fashioned) individual rights protections, and being met with demoralized and objectifying behavior of decision makers. This is not to say that modern AI does not deserve such scrutiny. With machine-learned ‘insights’ of increasing influence on facts, norms, and concepts used in laws, and in medical and other expert knowledge, informing all kinds of policies, one can certainly imagine a role for explanation regulation: after all, these are norms that express societal decisions of what is interest to know about a decisional process. But the need was already there and so, it seems our norms need re-imagining themselves. Abstracting from the EU’s ‘automation obsession’ also acknowledges that it is increasingly illusionary and therewith unproductive to distinguish manual and automated phases of decision making.

A path of investigations was set out accordingly. The thesis pursued understanding about relations between knowledge, the well-being of people, responsibilities of explainers, and the role of legal explanation rules. This critical analysis of law required a temporary (and inevitably artificial) suspension of committed and nuanced understanding of the system and a distancing from the type of legal research that takes it as its starting point. In a chapter that sources insights from the philosophical fields of epistemic justice and injustice, two first steps were taken. In how these fields of research investigate the (moral, ethical, instrumental, and theoretical) rights and (historical and contemporary) wrongs of knowledge practices, they meaningfully informed a re-idealized set of explanation duties.

To formulate this kind of guidance for ‘explanation,’ the interaction between explainers and explainees was cast as a practice of knowledge making about knowledge making (in law, expertise, and methods). Explanation rules, in such a view, govern conduct about conduct. Finally, the explainer-explainee relation was described as an interactive, testimonial practice that requires clear and publicly known rules of engagement. In honor of how critical theorists have warned to ‘start from the trouble’ in dealing with suspect ideals, insights from the fields were parsed in three dimensions (the misuse of epistemic authority (and the effects on people as a consequence), the perpetuation of wrongs in shared knowledge spheres, and the institutional promotion of preventative and corrective labor) and applied to model an epistemic justice oriented, epistemic injustice informed set of obligations for explainers. The modeled values describe ‘duties of explanation care’ related to four phases of an explanation cycle. Explainers are addressed as investigators of their own and their explainees’ information positions with

regard to their decisional paradigm (phases 1&2), as co-creators and communicators of knowledge in the form of decisions and their justifications (phase 3), and as reporters on decision and explanation processes (phase 4). Throughout the cycle, the focus of the duties is on surfacing possibly oppressive dimensions of decision paradigms (in underlying knowledges, decisional methods, and interaction.)

The application means to contribute to the growing recognition of the fields' relevance for work on justice-related aspects of data technologies and artificial intelligence. By showcasing a bespoke application for it in the context of explanation, the application is also a response to a much-heard argument in defense of inscrutable ADS methods, namely that inter-human understanding is an overrated commodity. The Model means to meaningfully sustain responsible explainers, as well as the critical assessment of how they are instructed: our legal explanation rules. Inter-human understanding has indeed been overrated in these rules, but in a different way. Epistemic in/justice theory shows how we are more versed in dealing with our 'black-boxedness' than some like us to think.

7.2 A tale of two domain studies

The second part of the thesis engaged the modeled epistemic in/justice values for a critical assessment of the legal explanation rules of two domains wherein the informational and social inequality of decision makers and explainees are typically substantive: administrative decisions (with a focus on dependent subjects) and General medical Practice or 'GP.' Such a study inevitably only renders an incomplete picture: just the contours of explanation governance as sketched in law. But by juxtaposing these contours with a knowledge-oriented description of the respective domains' decision making and decision makers, the incompleteness of law's guidance was made to speak and lost some of its innocence. Meaning, it is true that law's functioning, including its further development typically benefits from the use of open norms and less-than-minute instructions and obligations. But the studied domains' legal guidance shows an interesting gap between acknowledged needs for 'inequality compensation' in the form of elaborate explanations and explanations of a different character than the codified norms express. Such acknowledgement was not just found in (critical) scholarship about the domains but also in principled legal, ethical and professional norms.

The General Principles of Proper Administration for example contain two intimately related norms that are much named as inspirational in AI explanation debates: diligent preparation of decisions and proper motivation. They were codified minimally and mostly unrelated. Explanations typically focus on justifying outcomes, not testifying to a proper process. The minimalism means their application is left up to a practice that was acknowledged to be lacking already in the law's own codification discussions. Very long-standing concerns exist about the lack of insightful, trustworthy

and understandable relations with decision subjects in a domain that lacks clear rules for evidence and evidential burdens, which raises questions about consecutive governments' appetite for progress. The need for explicit guidance also shows in these codification discussions themselves, and in scholarly, political and judicial debates after the Benefits Scandal: there is still fundamental disagreement about their meaning and force. In addition to this, a substantive amount of legal intricacies pose obstacles for explainers who would want to go the extra mile. This is especially true for explainers of the primary decision, since more comprehensive explanation needs are reserved for the review phase rules and in practice. The paradigm relies on self-sufficient, bureaucratically, legally and otherwise literate and capable decision subjects to fight for rights they do not know they have. Since this unfair burden on explainees has worsened in digital times, it is concerning that the starting point reproduces in technology-focused explanation regulation.

Since Health Law leaves much norm setting up to the ethical and professional medical fields, the inclusion of some of this literature in the study of the GP domain helped to reveal the gaps in law's explanation related guidance. The social and political nature of medical knowledge, the domain's relatively recent move (and decidedly bumpy ride) away from purely authoritarian decision making, and the daunting task of responsibly sharing decisions with patients are well established in non-legal norms but compliance is reported to fall short. One can also question the undemocratic process through which such norms are typically established, which increases the dependence on personal attitudes and for example makes for highly different ethical understandings of what non-oppressive patient consent turns on. In other words, the domain deserves more. Especially since here as well, several choices that law does make become unhelpful beacons (such as when it deprioritizes 'non-instrumental' informational exchanges), or invites doctors to ignore the rules altogether (on the ground that these are 'disconnected' from actual explanation practice.)

A historical exploration showed how the historical absence of explanation obligations had profound adverse consequences for the quality of medical knowledge development and the fair and safe treatment of patients, a finding that resonated strongly with developments of and in the AI fields. Honest, mutually well-informed relations between decision makers in, and out of the know don't 'magically' establish in a world that is rife with unequal power relations. E.g., the persistent 'excuse' for not establishing such relations, described as the 'technological complexity argument' and used to obfuscate deficiencies in doctors' own capabilities and ignore the politics of so-called 'technical' medical knowledge is used quite literally in both fields. Perhaps, this explains why the studied ethical, professional and education norms stop short of recognizing the need to discuss historical and contemporary wrongs of medicine, or the sociality of medical knowledge in general. Such conversations are important for groups of patients whose trust is either understandably low or the opposite, too high, and this is a finding that also applies to the Administrative domain.

The domains' legal approaches are alike as well in their legal modeling of explainees. Dutch health law casts the doctor-patient relationship as one of contract, ignoring their highly unequal social and epistemic positions and patients' inherent vulnerability. Within this frame patients are cast as capable, self-managing, autonomous individuals. In both domains, knowledge related wrongs—alive in decisional aims, conceptualizations, and methods—are generally not acknowledged as of interest for legal explanation governance. Generally speaking, both sets of explanation rules do not instruct explainers to reach sufficiently meaningful information positions, nor to engage in the kind of interaction that allows to fundamentally respect explainees as knowers and rights holders. In how explainers are not instructed to see, understand, and cater to patients in relation to each other and to unfair societal structures, the two prototypical legal explanation paradigms share concerns that apply to the human rights framework that rules them in turn.

The right to explanation is not the only foundational legal protection that lacks force in face of AI's reproduction and exacerbation of oppressive power structures. Anti-discrimination law is another. Ant this paradigm too lacks force against 'manually' administered injustice. To improve the situation, Moyn has argued that the instead of requiring from those in power to act "more humane," Human Rights Law should demand a more just distribution of powers. The next section adds considerations on how to use the momentum that the proliferation of AI-informed wrongs is seen to have created. Again, the Benefits Scandal is a poster child for the need and for an integrated approach. When the institutional character of the wrongs were finally acknowledged, the State Secretary denied victims a claim for damages that the wrongful treatment had caused them. Institutional racism, he stated, lacks legal significance. The remark testifies to the depth of the problem addressed in the thesis.

7.3 An argument for care-ful progress of present and future explanation regulation

The ambition of the thesis is to provide something more than impressionistic guidelines, but something less than sufficient guidance. The 'modeled duties of explanation care' are meant to further all of our thinking about the point and governance of explanation. It is about time, but also, high time that this happens. The distance between codified, minimalistic explanation norms and acknowledged epistemic justice-oriented explanation needs in research and in ethical and professional norms can be seen to grow in AI-infused times. Not because law is getting worse, but because revelations of AI-informed wrongs are inspiring progressive efforts in other norm setting domains. In the Netherlands, the Benefits Scandal and several other revelations of institutional wrongdoing (e.g., in the police force) creates additional momentum that could be used to point out how such progress needs to stretch beyond the digitalized sphere. The thesis pointed out several ways in which this momentum can be made to work or,

adversely, how such work stands to be undermined, and how the modeled values of explanation care help to bring this to the fore.

The thesis for example engaged with progressive efforts in the sphere of AI and human rights impact and risk assessments that include explanation in their focus. Such assessments are on the brink of being broadly prescribed as a pre-employment and/or periodic, continuous condition for using ADS. They are still multiform, and approaches and methodologies for them are being developed widely. The processes inevitably engage with the fairness of knowledge that goes into, and comes out of decision making, and some usefully identify the need for a holistic (analog-digital) investigatory approach to decision spheres of deployment in order to do so. Efforts are made to create meaningful ‘explanatory’ knowledge through the assessment processes in which interested and affected parties and experts of various disciplines are brought together, and their mutual understanding and decision making is aimed for.

The domain analysis already showed how all parties will need guidance and training in order to make this work. The kinds of investigative and honest conversations that need to be had are not easy and there are few legal traditions in the context of decision making and explanation to guide them, which brings in the aforementioned gap. E.g., the medical domain showed how meaningful justificatory conversations with patients don’t tend to establish without strong legal intervention, but law’s reluctance to clearly treat explanation as a process and to include a focus on justifying ‘input knowledge’ has also been unhelpful here. The openness, assessability, and archival demands of the processes are important to note as well. The Administrative domain showed how an unhelpful dichotomy between negatively framed formal, ‘legalistic’ processes and positively framed informal ‘humaneness’ naturally lead to ungoverned spaces where power inequalities go unchecked.

Another concern is that what are (or at least may become) very useful processes in themselves don’t meaningfully integrate with the existing decision-making context and how it is legally governed. This can happen for various reasons. For one, there was no shortage of evidence in both domains that explainees were not served well enough, and useful ideas circulated in research and practice. Yet law’s uptake has been minimal. It is unclear why things would be different with regard to new evidence. The relation of assessment processes with (democratic) law and (less democratic) policy making processes is yet unclear. Examples of canceled wrongful algorithmic systems built on not-canceled wrongful law and policy were given. The reverse is problematic too: insufficient assessments should not get to legitimize wrongful decision making, same as policy rules have been used to avoid necessary consideration and justificatory reasoning.

This connects to the risk that underlying explanation paradigms will ‘corrupt’ more useful, digitally oriented governance when the two remain as separated as they are now. The rules in place still govern the manual phases of decision making and are therewith

of influence on the mixed-methods approaches that most ‘automated’ decision making now comprises of. This influence may even grow when the large amount of work that it takes to truly assess ADS is not supported with necessary financial and organizational means. There are already tendencies to restrictively categorize which ‘algorithmic applications’ should be considered risky enough to assess, for example by following the risk categorization of the proposed EU AI Act despite much criticism about how this is inadequate. Another way to restrict the necessary investments and additionally avoid to have to explain ‘technology’ is to focus on justification of outcomes over legitimizing processes, which further limits the causal understanding that the modeled values strongly advocate. And in light of the time investments that are needed to study the interplay of qualitative, quantitative, and automated methods at use in the assessed decisional process, the non-digital dimensions risk to be skipped. The same is true for connecting with the right persons in assessment: they need to be found and may be unwilling to connect on the basis of bad experience in the decision domain. Ironically, when this (rightly) leads to decisions to abort an ADS design, the progressive decision assessment ambitions stand to be aborted with them.

But identified flaws also come in via the reliance on review procedures and idealizations of ‘the human in the loop’ as is done and suggested in and for ADM explanation governance. Administrative law principles and medical professional ethics are not just named as inspiration but also as fallback options for ‘gaps’ in ADM regulation. In such cases, all identified concerns apply. Human explainers cannot simply be relied on to have either the capabilities, the inclination, or the organizational and legal ‘blessing’ to engage with injustices in their underlying knowledges and decisional methods, and investigate whether explainees were able to participate responsibly. At the same time, care needs to be taken that blaming humans in these positions alone for ‘biased’ outcomes diverts necessary attention from the larger system in which they are expected to function, and the problematic political influence of these system’s roots. Such diverting for example expresses in the Dutch state’s defense of a broad ‘simple automation for public good’ exception clause from the General Data Protection Regulation’s ‘right to explanation’ trigger provision that forbids such processing unless precautions are taken: Article 22. The State reasoned that gravely unjust outcomes from notorious ‘simple’ cases that tend to befallen large groups of less privileged decision subjects, such as automated fines based on obtuse registrations in unkempt registers, were down to the civil servant body at the post-processing end of the decisional chain. But when we want such explainers to take moral responsibility, and we do, the State needs to make sure they are in a position to do so.

Care to explain?

Bibliography

- ‘A survey of artificial intelligence risk assessment methodologies: The global state of play and leading practices identified’. EY and Trilateral Research, 2021.
- Achbab, Samir. ‘De Toeslagenaffaire is ontstaan uit institutioneel racisme’. *NRC*. Consulted 1 November 2021. <https://www.nrc.nl/nieuws/2021/05/30/de-toeslagenaffaire-is-ontstaan-uit-institutioneel-racisme-a4045412>.
- AI HLEG. ‘Ethics Guidelines for Trustworthy AI’. Text. European Commission, 8 April 2019. <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.
- ‘AI Is Wrestling with a Replication Crisis’. *MIT Technology Review*. Consulted 13 November 2020. <https://www.technologyreview.com/2020/11/12/1011944/artificial-intelligence-replication-crisis-science-big-tech-google-deepmind-facebook-openai/>.
- Alcoff, Linda Martin. ‘Philosophy and Philosophical Practice: Eurocentrism as an Epistemology of Ignorance’. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- ‘Algemene bepalingen van administratief recht, 5e herziene edition’. Commissie ABAR, 1984.
- ‘Algorithmic Impact Assessment: A Case Study in Healthcare’. Ada Lovelace Institute, 8 February 2022.
- AlgorithmWatch. ‘AlgorithmWatch’s Response to the European Commission’s Proposed Regulation on Artificial Intelligence – A Major Step with Major Gaps’, April 2021. <https://algorithmwatch.org/en/response-to-eu-ai-regulation-proposal-2021>.
- Allen, Amy. ‘Power/Knowledge/Resistance: Foucault and Epistemic Injustice’. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Allen, Garland E. ‘The Ideology of Elimination: American and German Eugenics, 1900-1945’. In *Medicine and Medical Ethics in Nazi Germany: Origins, Practices, Legacies*, edited by Francis R. Nicosia and Jonathan Huener, 1st Edition. New York: Berghahn Books, 2002.
- Amann, Julia, Alessandro Blasimme, Effy Vayena, Dietmar Frey, Vince I. Madai, and the Precise4Q consortium. ‘Explainability for artificial intelligence in healthcare: a multidisciplinary perspective’. *BMC Medical Informatics and Decision Making* 20, nr. 1 (30 November 2020): 310.
- ‘Amended proposal for a Council Directive on the protection of individuals with regard to the processing of personal data and on the free movement of such data’. Commission of the European Communities, 15 October 1992.
- Amnesty International. ‘Aangifte bedreiging en vernieling Sint-intocht Staphorst’, 21 November 2022. <https://www.amnesty.nl/actueel/amnesty-nederland-doet-aangifte-van-bedreiging-rond-demonstratie-in-staphorst>.
- Amoore, Louise. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Duke University Press, 2020.
- Ananny, Mike, and Kate Crawford. ‘Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability’. *New Media & Society* 20, nr. 3 (1 March 2018): 973–89.
- Anderson, Elizabeth. ‘Epistemic Justice as a Virtue of Social Institutions’. *Social Epistemology* 26, nr. 2 (1 April 2012): 163–73.

- Aquino, Yves Saint James, and Stacy M. Carter. 'Explanation versus Outcome: Examining Professional Perspectives on the Ethics of Explainable AI in Clinical Diagnosis', 2022. <https://juanmduran.net/explanation-versus-outcome-examining-professional-perspectives-on-the-ethics-of-explainable-ai-in-clinical-diagnosis/>.
- Arendt, Hannah. *Responsibility and Judgment*. Reprint edition. Schocken, 2005.
- . 'Some questions of Moral Philosophy'. In *Responsibility and Judgment*, Reprint edition. Schocken, 2005.
- Article 29 Working Party. 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (WP 251rev.1)'. European Commission, 6 February 2018.
- . 'Guidelines on Transparency under Regulation 2016/679 (wp260rev.01)'. European Commission, 11 April 2018.
- Asghari, Hadi, Birner, Nadine, Burchardt, Aljoscha, Daniela Dicks, Faßbender, Judith, Feldhus, Nils, Hewett, Freya, et al 'What to Explain When Explaining Is Difficult. An Interdisciplinary Primer on XAI and Meaningful Information in Automated Decision-Making'. Zenodo, 22 March 2022. <https://doi.org/10.5281/ZENODO.6375784>.
- 'Automated Decision-Making Under the GDPR - A Comprehensive Case-Law Analysis'. Future of Privacy Forum, May 2022.
- Baalen, Sophie Jacobine van. 'Knowing in Medical Practice: Expertise, Imaging Technologies and Interdisciplinarity', 2019. <https://research.utwente.nl/en/publications/knowning-in-medical-practice-expertise-imaging-technologies-and-in>.
- Baalen, Sophie Jacobine van, en Mieke Boon. 'Evidence-Based Medicine versus Expertise: Knowledge, Skills and Epistemic Actions'. *Knowing and Acting in Medicine*, 2017, 21–38.
- Baalen, Sophie van, and Mieke Boon. 'An Epistemological Shift: From Evidence-Based Medicine to Epistemological Responsibility'. *Journal of Evaluation in Clinical Practice* 21, nr. 3 (2015): 433–39.
- Baalen, Sophie van, and Annamaria Carusi. 'Implicit Trust in Clinical Decision-Making by Multidisciplinary Teams'. *Synthese* 196, nr. 11 (1 November 2019): 4469–92.
- Bacchi, Carol. 'Why Study Problematizations? Making Politics Visible'. *Open Journal of Political Science* 2, nr. 1 (26 April 2012): 1–8.
- Bakker, Rob. *Boekhouders van de Holocaust: Nederlandse ambtenaren en de collaboratie*. Holocaust bibliotheek. Hilversum: uitgeverij Verbum, 2020.
- Balayn, Agathe, and Seda Gürses. 'Beyond-Debiasing: Regulating AI and its inequalities'. EDRi, 2021.
- Bar-Itzhak, Chen, Micol Bez, Angelo Vannini, and Victoria Zurita. 'In Search of Epistemic Justice: A Tentative Cartography'. *University of Pennsylvania international CFP listing* (blog). Consulted 13 December 2021. <https://call-for-papers.sas.upenn.edu/cfp/2021/12/09/in-search-of-epistemic-justice-a-tentative-cartography>.
- Barocas, Solon, and Andrew D. Selbst. 'Big Data's Disparate Impact'. *California Law Review* 104, nr. 3 (2016): 671–732.
- Battaly, Heather. 'Testimonial Injustice, Epistemic Vice, and Vice Epistemology'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.

- Bayamlioğlu, Emre. 'The Right to Contest Automated Decisions under the General Data Protection Regulation: Beyond the so-Called "Right to Explanation"'. *Regulation & Governance* 16, nr. 4 (2022): 1058–78.
- Beckers, Marion, Toine De Bie, Dolf Van Harinxma thoe Slooten, René Klomp, en Anneke Rang, red. *Motiveren, Over het motiveren van rechterlijke uitspraken*. Prinsengrachtreeks. Ars Aequi Libri, 2017.
- Beckwith, J., and R. Pierce. 'Genes and Human Behavior: Ethical Implications.' In *Molecular-Genetic and Statistical Techniques for Behavioral and Neural Research*, edited by RT Gerlai. Elsevier Academic Press, 2018.
- Bell, Derrick A. 'Who's Afraid of Critical Race Theory? RT_1995UIILLRev893.pdf'. *University of Illinois law Review* 1995, nr. 893 (1995).
- Benjamin, Ruha. *Race After Technology*. Polity Press, 2019.
- Berg, Marc, and Annemarie Mol, red. *Differences in Medicine: Unraveling Practices, Techniques, and Bodies*. Body, Commodity, Text. Duke University Press, 1998.
- Berge, Lukas van den. 'Bestuursrecht na de toeslagenaffaire: hoe nu verder? Over het rechtskarakter van het bestuursrecht'. *Ars Aequi* 2021 (November 2021): 987.
- Besluit tot boeteoplegging Minister van Financien (Autoriteit Persoonsgegevens 25 November 2021).
- Besselink, Leonard. 'De Afdeling Bestuursrechtspraak en de rechtsstatelijke crisis van de Toeslagenaffaire'. *NJB* 2021, nr. 3.
- Bhatt, Umang, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José M. F. Moura, and Peter Eckersley. 'Explainable machine learning in deployment'. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 648–57. FAT* '20. Barcelona, Spain: Association for Computing Machinery, 2020. <https://doi.org/10.1145/3351095.3375624>.
- Binns, Reuben. 'Human Judgment in Algorithmic Loops: Individual Justice and Automated Decision-Making'. *Regulation & Governance* 16, nr. 1 (2022): 197–211.
- Binns, Reuben, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. "'It's Reducing a Human Being to a Percentage"; Perceptions of Justice in Algorithmic Decisions'. In *Proceedings of the 2018CHI Conference on Human Factors in Computing Systems*, 1–14, 2018.
- Binns, Reuben, and Michael Veale. 'Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR'. *International Data Privacy Law* 11, nr. 4 (2021).
- Birkinshaw, Patrick. 'Transparency as a Human Right'. In *Transparency: The Key to Better Governance?*, edited by Christopher Hood and David Heald. Oxford, UK: Oxford University Press, 2011.
- Black, Edwin. *IBM and the Holocaust: The Strategic Alliance Between Nazi Germany and America's Most Powerful Corporation*. Crown Publishing Group, 2001.
- Black, Monica. *A Demon-Haunted Land: Witches, Wonder Doctors, and the Ghosts of the Past in Post-WWII Germany*. New York, New York: Metropolitan Books, 2020.
- Booy, Heleen, en Kiki Varenkamp. 'Diversiteit als toetje: het filosofie-examen gaat over 44 mannen, vier vrouwen en één filosoof van kleur'. *De Groene Amsterdammer*, 2021.

- Borghans, L., en R.E.M. Dieris. 'Ongelijkheid in het nederlandse onderwijs door de jaren heen'. In *Preadviezen voor de Koninklijke Vereniging voor Staathuishoudkunde*, edited by A. Gielen, D. Webbink, en B. ter Weel. ESB & Koninklijke Vereniging voor Staathuishoudkunde, 2021.
- Borst, Wim. 'Mag het bestuur ook wat de rechter mag? Over de verhouding tussen bestuur en rechter (naar aanleiding van de toeslagenaffaire)'. *Ars Aequi* 2022, nr. April .
- Bos, René ten. *Bureaucratie is een inktvis*. Amsterdam: Boom Uitgevers, 2015.
- Bot, Michiel. 'Is institutioneel racisme echt racistisch?'. *NJB* 26, 2022.
- Bovens, Mark. 'Analysing and Assessing Accountability: A Conceptual Framework1'. *European Law Journal* 13, nr. 4 (2007): 447–68.
- Bowker, Geoffrey C., en Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. Revised edition. Cambridge, Massachusetts London, England: The MIT Press, 2000.
- Brandon, Pepijn, Guno Jones, Nancy Jouwe, en Matthias Van Rossum, red. *De slavernij in Oost en West: het Amsterdam onderzoek*, 2021.
- Brennan-Marquez, Kiel. 'Plausible Cause: Explanatory Standards in the Age of Powerful Machines'. *Vanderbilt LawReview* 70 (2017): 1249–1301.
- Brenninkmeijer, Alex. 'De burger tussen de ambities en doelstellingen van de Awb'. In *25 jaar Awb: in eenheid en verscheidenheid*, edited by A.T. Marseille, Tom Barkhuysen, Willemien den Ouden, Hans Peters, en Raymond Schlössels. Deventer: Wolters Kluwer, 2019.
- . 'Welke lessen zijn te trekken uit de kinderopvangoeslagaffaire en de problemen bij uitvoeringsorganisaties?'. In *Grensoverstijgende rechtsbeoefening: Liber amicorum Jan Jans*, edited by K. J. de Graaf, Bert Marseille, Sacha Prechal, Rob Widdershoven, en Heinrich Winter. Zutphen: Paris, 2021.
- Bridges, Khiara M. *The Poverty of Privacy Rights*. Stanford University Press, 2017.
- 'Brief van de ministers voor Rechtsbescherming en van Binnenlandse Zaken en Koninkrijksrelaties'. Parliamentary papers 34 775 VI, no 4, 22 January 2018.
- Bröring, Herman, en Albertjan Tollenaar. 'Menselijke maat in het bestuursrecht: afwijken van algemene regels'. In *Grensoverstijgende rechtsbeoefening: Liber amicorum Jan Jans*, edited by K. J. de Graaf, Bert Marseille, Sacha Prechal, Rob Widdershoven, en Heinrich Winter. Zutphen: Paris, 2021.
- Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World*. The MIT Press, 2018.
- Brouwers, Lucas, en Patricia Veldhuis. 'Robbert Dijkgraaf: "Politici hebben heel besliste meningen. Daar moet ik aan wennen"'. *NRC*, 11 March 2022. <https://www.nrc.nl/nieuws/2022/03/11/robert-dijkgraaf-politici-hebben-heel-besliste-meningen-daar-moet-ik-aan-wennen-a4100635>.
- Brownword, Roger. 'In the year 2061: from law to technological management'. *Law, Innovation and Technology* 7, nr. 1 (2 January 2015): 1–51.
- . 'Law Disrupted, Law Re-Imagined, Law Re-Invented'. *Technology and Regulation*, 20 May 2019, 10–30.
- Bryson, Joanna. 'The origins of bias and the limits of transparency', Presentation at Dagstuhl 21231, 9 June 2021.

- ‘Buitenhof’. *Buitenhof*, 12 April 2015. <https://tvblik.nl/buitenhof/12-April-2015>.
- Burgers, J. ‘Oratie: Persoonsgerichte zorg en richtlijnen: contradictie of paradox?’ Oratie. Maastricht, 2017. <https://cris.maastrichtuniversity.nl/en/publications/persoonsgerichte-zorg-en-richtlijnen-contradictie-of-paradox>.
- Butler, Judith. *The Force of Nonviolence: An Ethico-Political Bind*. Verso Books, 2021.
- Bygrave, Lee A. ‘Machine Learning, Cognitive Sovereignty and Data Protection Rights with Respect to Automated Decisions’. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 29 October 2020. <https://papers.ssrn.com/abstract=3721118>.
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. ‘Semantics derived automatically from language corpora contain human-like biases’. *Science* 356, nr. 6334 (14 April 2017): 183–86.
- Calo, Ryan. ‘The Scale and the Reactor’. SSRN Scholarly Paper. Rochester, NY, 9 April 2022. <https://doi.org/10.2139/ssrn.4079851>.
- Calo, Ryan, and Danielle K Citron. ‘The Automated Administrative State: A Crisis of Legitimacy’. *Emory Law Journal* 70, nr. 4 (2021).
- Çankaya, Sinan. ‘Opinie | Ze bedoelden het wél zo – het racisme kan onmogelijk ontkend worden’. *NRC*, 27 May 2022. <https://www.nrc.nl/nieuws/2022/05/27/ze-bedoelden-het-wel-zo-het-racisme-kan-onmogelijk-ontkend-worden-a4129407>.
- Carel, Havi, and Ian James Kidd. ‘Epistemic Injustice in Medicine and Health Care’. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Carolan, Eoin. ‘The Continuing Problems with Online Consent under the EU’s Emerging Data Protection Principles’. *Computer Law & Security Review* 32, nr. 3 (June 2016): 462–73.
- Carson, Ronald A. ‘Medical Ethics as Reflective Practice’. In *Philosophy of Medicine and Bioethics: A Twenty-Year Retrospective and Critical Appraisal*, edited by Ronald A. Carson en Chester R. Burns, 181–91. Philosophy and Medicine. Dordrecht: Springer Netherlands, 1997.
- Case 19-1392 Dobbs v. Jackson Women’s Health Organization (06/24/2022) (US Supreme Court 2022).
- Chouldechova, Alexandra. ‘Fair prediction with disparate impact: A study of bias in recidivism prediction instruments’. *arXiv:1703.00056 [cs, stat]*, 28 February 2017. <http://arxiv.org/abs/1703.00056>.
- Citron, Danielle Keats. ‘Technological Due Process’. *Washington University Law Review* 85, nr. 6.
- Citron, Danielle Keats, en Frank A. Pasquale. ‘The Scored Society: Due Process for Automated Predictions’. *University of Maryland Francis King Carey School of Law Legal Studies Research Paper* 2014, nr. 8 (2014).
- Claessens, Machteld. ‘Het (on)nut van een recht op toegang tot de overheid als nieuw algemeen beginsel van behoorlijk bestuur’. *Nederlands Tijdschrift voor Bestuursrecht* 2021/105, nr. 4 (April 2021).
- ‘Coalitieakkoord 2021 – 2025 Omzien naar elkaar, vooruitkijken naar de toekomst’. VVD, D66, CDA en ChristenUnie, 15 December 2021.

- Cobb, Matthew. 'Why Your Brain Is Not a Computer'. *The Guardian*, 27 February 2020, sec. Science. <https://www.theguardian.com/science/2020/feb/27/why-your-brain-is-not-a-computer-neuroscience-neural-networks-consciousness>.
- Codagnone, C, and G Liva. 'Identification and Assessment of Existing and Draft Legislation in the Digital Field'. *EU Policy Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies*, , 82.
- Code, Lorraine. *Epistemic Responsibility*. Brown University Press, 1987.
- . 'Epistemic Responsibility (2017)'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Coglianese, Cary, and David Lehr. 'Transparency and Algorithmic Governance'. SSRN Scholarly Paper. Rochester, NY, 9 November 2018. <https://papers.ssrn.com/abstract=3293008>.
- . 'Transparency and Algorithmic Governance'. *Administrative Law Review* 71, nr. P.1 (2019).
- Cohen, Felix. 'The Ethical Basis of Legal Criticism'. *The Yale Law Journal* 41, nr. 2 (1931): 201–20.
- Colaner, Nathan. 'Is Explainable Artificial Intelligence Intrinsically Valuable?' *AI & Society* 37, nr. 1 (1 March 2022): 231–38.
- Congdon, Matthew. 'What's Wrong with Epistemic Injustice? Harm, Vice, Objectification, Misrecognition'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Conway, Martin. 'The Certainties of the Past? The Making of Democracy in Western Europe after 1945'. Amsterdam, 26 January 2023. <https://spui25.nl/programma/the-certainties-of-the-past>.
- Coulter, A., V. Entwistle, and D. Gilbert. 'Sharing Decisions with Patients: Is the Information Good Enough?' *BMJ* 318, nr. 7179 (30 January 1999): 318–22.
- Council of Europe. *The Administration and You, A Handbook: Principles of administrative law concerning relations between individuals and public authorities*. Council of Europe Publishing, 2018.
- Crenshaw, Kimberle. 'Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color'. *Stanford Law Review* 43, nr. 6 (1991): 1241–99.
- Dalla Corte, Lorenzo. 'A Right to a Rule: On the Substance and Essence of the Fundamental Right to Personal Data Protection'. In *Data Protection and Privacy: Data Protection and Democracy*, edited by D. Halliman, R. E. Leenes, S. Gutwirth, and P. De Hert. Hart Publishing, 2020.
- Damen, Leo. 'De autonome Awbmen?' *Ars Aequi* 66, nr. 07–08 (2017).
- . 'Ik was het niet, ik was het niet, het was de wetgever!', , 4.
- Dartel, Hans van, en Bert Molenwijk, red. *In gesprek blijven over goede zorg: Overlegmethoden voor moreel beraad*. Boom Filosofie, 2014.
- David Graeber. *The Utopia of Rules: On Technology, Stupidity, and the Secret Joys of Bureaucracy*. Melville House Books, 2015.

- De blauwe familie*, 2022. <https://www.2doc.nl/documentaires/2022/05/de-blauwe-familie.html>.
- De Boer, Marjolijn, en Sylvana Van den Braak. 'Verhuizen om wèl hulp te krijgen?' *De Groene Amsterdammer*, 25 September 2019.
- 'De burger kan niet wachten: Jaarverslag van de Nationale ombudsman, de Kinderombudsman en de Veteranenombudsman over 2021'. Nationale Ombudsman, 2022.
- De Graaf, K J, en A T Marseille. 'Exit willekeurstoets. Bestuursrechterlijke toetsing aan het evenredigheidsbeginsel na 2-2-'22'. *Ars Aequi* 2022, nr. April .
- De Jong, Frits. "'Er is een Wabo nodig voor het sociaal domein" | iBestuur'. Consulted 17 January 2019. <https://ibestuur.nl/partner-vng-realisatie/er-is-een-wabo-nodig-voor-het-sociaal-domein>.
- De Koster, Yolanda. 'Wisselend succes na mediation sociaal domein'. *Binnenlands Bestuur* (blog), 16 November 2018. <https://www.binnenlandsbestuur.nl/sociaal/nieuws/wisselend-succes-na-mediation-sociaal-domein.9601304.lynx>.
- 'De Monitor: Enorme boetes bij onverzekerd rijden'. *De Monitor*, 2015. <https://kro-ncrv.nl/persberichten/de-monitor-enorme-boetes-bij-onverzekerd-rijden>.
- De Nationale Ombudsman. 'Burgerperspectief: een manier van kijken'. Jaarverslag 2015. Nationale Ombudsman, 2015.
- De Ruijter, Wouter, Aart Hendriks, en Marian Verkerk, red. *Huisarts tussen individu en familie: morele dilemma's in de huisartspraktijk*. Van Gorcum, 2012.
- Dehue, Trudy. 'Definities die oorzaken worden'. In *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen. Boom Filosofie, 2013.
- Desai, Deven R., and Joshua A. Kroll. 'Trust But Verify: A Guide to Algorithms and the Law'. *Harvard Journal of Law & Technology* 31, nr. 1 (27 April 2017).
- 'Digitale dokters: Een ethische verkenning van medische expertsystemen - Signalement - CEG – Centrum voor Ethiek en Gezondheid'. Centrum voor Ethiek en Gezondheid, 2018.
- 'Digitalisering: wetgeving en bestuursrechtspraak'. Raad van State, May 2021.
- Doshi-Velez, Finale, Mason Kortz, Ryan Budish, Christopher Bavitz, Samuel J. Gershman, David O'Brien, Stuart Shieber, Jim Waldo, David Weinberger, en Alexandra Wood. 'Accountability of AI Under the Law: The Role of Explanation'. Berkman Center Research Publication, forthcoming, November 2017.
- Dotson, Kristie. 'Conceptualizing Epistemic Oppression'. *Social Epistemology* 28, nr. 2 (3 April 2014): 115–38.
- Dusenbery, Maya. *Doing Harm: The Truth About How Bad Medicine and Lazy Science Leave Women Dismissed, Misdiagnosed, and Sick*. HarperCollins, 2018.
- Dwarswaard, Jolanda. 'De Dokter en de Tijdgeest'. Erasmus university, 2011.
- Eck, Marlies van. 'Geautomatiseerde ketenbesluiten & rechtsbescherming: Een onderzoek naar de praktijk van geautomatiseerde ketenbesluiten over een financieel belang in relatie tot rechtsbescherming.' Doctoral Thesis, Tilburg University, 2018.
- ECLI:NL:CRVB:2014:1035, Centrale Raad van Beroep, 13-4228 WWB, No. ECLI:NL:CRVB:2014:1035 (CRvB 20 March 2014).
- ECLI:NL:HR:1987:AG5500, voorheen LJN AG5500, AC0705, AJ3785, AM9322, Hoge Raad, 12.717, No. ECLI:NL:HR:1987:AG5500 (HR 9 January 1987).

- ECLI:NL:HR:2021:995, Hoge Raad, 19/03033, No. ECLI:NL:HR:2021:995 (Hoge Raad 25 June 2021).
- ECLI:NL:RBAMS:2022:3066 Geautomatiseerde afwijzing na invullen online vragenlijst is bestuursrechtelijk besluit (Rechtbank Amsterdam October 2022).
- EDPS Ethics Advisory Group. 'Towards a digital ethics', 2018.
- Edwards, Lilian, en Michael Veale. 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' *IEEE Security & Privacy* 018, nr. 16(3) (2018).
- . 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For'. *Duke Law & Technology Review* 16, nr. 18 (23 May 2017).
- 'Een ongeluk komt nooit alleen: Een rapportage over een geslaagde interventie van de Nationale ombudsman naar aanleiding van een klacht over het Centraal Justitieel Incassobureau (CJIB) te Leeuwarden en de Dienst Wegverkeer (RDW) te Zoetermeer'. Nationale Ombudsman, 13 January 2015.
- Nieuwssite BZK. 'Eerste stap naar constitutionele toetsing door de rechter gezet'. Consulted 8 July 2022. <https://www.nieuwsbzk.nl/2253818.aspx>.
- 'Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements'. Feldman Barret et al., Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements', *Psychological Science in the Public Interest* volume 20, issue 1, 2019
- Engelberts, Ingeborg, Maartje Schermer, en Awee Prins. 'Een goed gesprek is de beste persoonsgerichte zorg'. *Medisch Contact* 2018, nr. 28/29. Consulted 16 July 2020. <https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/een-goed-gesprek-is-de-beste-persoonsgerichte-zorg.htm>.
- Essers, Geurt. 'De huisartsopleiding in Nederland is al 50 jaar een succesformule'. *Huisarts en wetenschap* 64, nr. 7 (July 2021): 6–8.
- Sociale Vraagstukken Sociale Vraagstukken: Wetenschappers & professionals over maatschappelijke kwesties. 'Ethici, even pas op de plaats', 27 March 2020. <https://www.socialevraagstukken.nl/ethici-even-pas-op-de-plaats/>.
- Ettekoen, B.J. van, en A.T Marseille. 'Afscheid van de klassieke procedure in het bestuursrecht?' In *Afscheid van de klassieke procedure?*, 2017de–1ste dr. Vol. 147. Handelingen Nederlandse Juristen-Vereeniging. Wolters Kluwer, 2017.
- Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. 1st edition. St. Martin's Press, Macmillan Publishers, 2018.
- European Commission. 'Stronger protection, new opportunities: Commission guidance on the direct application of the General Data Protection Regulation as of 25 May 2018, COM(2018) 43 final', 24 January 2018.
- 'Explanatory text for Proposal for a Council Directive concerning the protection of individuals in relation to the processing of personal data'. Commission of the European Communities, 1990.
- Faden, Ruth R., and Tom L. Beauchamp. *A History and Theory of Informed Consent*. New York: Oxford University Press, 1986.
- Fenster, Mark. 'Transparency in Search of a Theory'. *European Journal of Social Theory* 18, nr. 2 (1 May 2015): 150–67.

- Feteris, E., and H. Kloosterhuis. 'The Analysis and Evaluation of Legal Argumentation: Approaches from Legal Theory and Argumentation Theory'. *Studies in Logic, Grammar and Rhetoric* 16 (2009).
- The Guardian. "'Fewer Rights than Their Grandmothers': Read Three Justices' Searing Abortion Dissent", 24 June 2022, sec. Opinion. <https://www.theguardian.com/commentisfree/2022/jun/24/supreme-court-roe-v-wade-breyer-sotomayor-kegan>.
- Filet, B.C. *Kortsluiting met de bureaucratie : over participatiemogelijkheden van burgers bij het openbaar bestuur*. Bestuur-Bestuurden, 1974.
- Fogteloo, Margreet. 'Hoogleraar mensenrechten Barbara Oomen: We zijn nonchalant over onze rechtsstaat'. *De Groene Amsterdammer*, 30 June 2021.
- Foster, Charles. *Human Dignity in Bioethics and Law*. Hart Publishing, 2011.
- Frankfurt, Harry D. *On Bullshit*. Princeton University Press, 2005.
- Franssen, P.E.M. 'Beleidsregels en de inherente afwijkingsbevoegdheid: de nieuwe lijn van de Afdeling.' *Praktisch Bestuursrecht* 2017, nr. 7 (13 December 2017).
- Fricker, Miranda. *Epistemic Injustice: Power and the Ethics of Knowing*. *Epistemic Injustice*. Oxford University Press, 2007.
- . 'Rational Authority and Social Power: Towards a Truly Social Epistemology'. *Proceedings of the Aristotelian Society, New Series* 98 (1988): 159–77.
- Friele, R D, J Legemaate, R P Wijne, R T Munshi, L J Knap, R J R Bouwman, en V D V Sankatsing. 'Wet kwaliteit, klachten en geschillen zorg', , 276.
- Friele, Roland, en Remco Coppen. 'Wetgeving en de positie van de patiënt: instrument voor verandering of terugvaloptie?' *Recht der Werkelijkheid*, 2010, 14.
- Fuller, Lon L. 'Positivism and Fidelity to Law: A Reply to Professor Hart'. *Harvard Law Review* 71, nr. 4 (1958): 630–72.
- Garvey, Shunryu Colin. 'Unsavoury Medicine for Technological Civilization: Introducing "Artificial Intelligence & Its Discontents"'. *Interdisciplinary Science Reviews* 46, nr. 1–2 (3 April 2021): 1–18.
- Gebru, Timnit, and Emily Denton. 'Tutorial on Fairness Accountability Transparency and Ethics in Computer Vision', 2020. <https://sites.google.com/view/fatecv-tutorial/home>.
- Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Dauméé III, and Kate Crawford. 'Datasheets for Datasets'. *ArXiv:1803.09010 [Cs]*, 23 March 2018. <http://arxiv.org/abs/1803.09010>.
- Geest, Michiel van der. 'Huisartsen ontsteld over vaccinatie-advies: "Niet verwacht dat Nederland hier zo klungelig mee om zou gaan"'. *de Volkskrant*, 12 April 2021, sec. Topverhalen vandaag. <https://www.volkskrant.nl/gs-bba4245c>.
- 'Gegijzeld door het Systeem. Onderzoek Nationale ombudsman over het gijzelen van mensen die boetes wel willen, maar niet kunnen betalen'. Nationale Ombudsman, 2015.
- Gerards, J. H. 'Meer rechtsbeginselen in de Awb? Gezichtspunten voor toekomstige codificatie'. In *15 jaar Awb*, edited by Tom Barkhuysen, Willemien den Ouden, en J.E.M. Polak. Boom Juridische Uitgevers, 2010.
- Gescinska, Alicja. *Kinderen van Apaté, Over leugens en waarachtigheid*. Lemniscaat, 2020.
- Geuskens, Machteld. 'Epistemic Justice: A Principled Approach to Knowledge Generation and Distribution'. Tilburg University, 2018.

- Gitelman, Lisa, red. *Raw Data Is an Oxymoron*. MIT press, 2013.
- Glas, Gerrit. 'Ziekte en stoornis in de psychiatrie'. In *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen. Boom Filosofie, 2013.
- Glenza, Jessica. 'How Dismantling Roe v Wade Could Imperil Other "Core, Basic Human Rights"'. *The Guardian*, 11 December 2021, sec. US news. <https://www.theguardian.com/us-news/2021/dec/11/supreme-court-roe-v-wade-gay-rights-contraceptives-fertility-treatments>.
- Glenza, Jessica, and Martin Pengelly. 'US Supreme Court Overturns Abortion Rights, Upending Roe v Wade'. *The Guardian*, 24 June 2022, sec. World news. <https://www.theguardian.com/world/2022/jun/24/roe-v-wade-overturned-abortion-summary-supreme-court>.
- Gopal, Priyamvada. 'On Decolonisation and the University'. *Textual Practice* 35, nr. 6 (3 June 2021): 873–99.
- Gordon, Lewis. *Freedom, Justice, and Decolonization*. Routledge, 2020.
- . 'Shifting the Geography of Reason'. Gepresenteerd bij Spinoza Lecture - University of Amsterdam, 24 May 2022. <https://www.uva.nl/en/shared-content/faculteiten/en/faculteit-der-geesteswetenschappen/events/events/2022/05/spinoza-1.html>.
- South Dakota State news website. 'Gov. Noem and Legislative Leaders Announce Plans for Special Session to Save Lives, Help Mothers'. Consulted 25 June 2022. <https://news.sd.gov/newsitem.aspx?id=30323>.
- Graaf, K. J. de, J H Jans, A T Marseille, and J de Ridder, red. *Quality of Decision-Making in Public Law: Studies in Administrative Decision-Making in the Netherlands*. Groningen: Europa Law Pub, 2007.
- Grasswick, Heidi. 'Epistemic Injustice in Science'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Green, Ben. 'The Flaws of Policies Requiring Human Oversight of Government Algorithms'. *Computer Law & Security Review* 45 (2022).
- Green, Ben, and Yiling Chen. 'Algorithmic Risk Assessments Can Alter Human Decision-Making Processes in High-Stakes Government Contexts'. *Proceedings of the ACM on Human-Computer Interaction* 5, nr. CSCW2 (18 October 2021): 418:1-418:33. <https://doi.org/10.1145/3479562>.
- Greene, Daniel, Anna Lauren Hoffmann, and Luke Stark. 'Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning', 2019. <https://doi.org/10.24251/HICSS.2019.258>.
- Greenfield, Adam. *Radical Technologies: The Design of Everyday Life*. London ; New York: Verso, 2017.
- Groot, Aviva de, and Sascha van Schendel. 'Explaining Responsibly: a panel discussion with Reuben Binns, Michael Veale, Martijn van Otterlo, and Rune Nyrop'. Tilburg, 2019. <https://easychair.org/smart-program/TILTING2019/2019-05-17.html#talk:89359>.

- Groot, Aviva de, Linnet Taylor, Merel Noorman, Siddhart De Souza, Gert Meyers, en Tineke Broer. 'Technologie is niet neutraal, dat zou Dijkgraaf moeten weten'. *ScienceGuide* (blog), 5 April 2022. <https://www.scienceguide.nl/2022/04/technologie-is-niet-neutraal-dat-zou-dijkgraaf-moeten-weten/>.
- Grootelaar, Hilke, en Kees van den Bos. 'De Awb vanuit een procedurele rechtvaardigheids-perspectief: hulpmiddel, hinderpaal of handvat? Macht en tegenmacht in de netwerksamenleving'. In *25 jaar Awb: in eenheid en verscheidenheid*, edited by A.T Marseille, Tom Barkhuysen, Willemien den Ouden, Hans Peters, en Raymond Schlössels. Deventer: Wolters Kluwer, 2019.
- Gürses, Seda, and Joris van Hoboken. 'Privacy after the Agile Turn (Version 3)'. *Open Science Framework*, 2017 <https://osf.io/preprints/socarxiv/9gy73/>.
- Haataja, Meeri, Linda van de Fliert, and Pasi Rautio. 'Public AI Registers: Realising AI transparency and civic participation in government use of AI Whitepaper written by Meeri Haataja, Linda van de Fliert and Pasi Rautio'. Gemeente Amsterdam. Last consulted 30 November 2022. <https://algoritmeregister.amsterdam.nl/wp-content/uploads/White-Paper.pdf>.
- Hafkenscheid, Dorothee. "'Mensen hebben het recht om te vragen om het AstraZeneca-vaccin'". *Medisch Contact* (blog). Consulted 9 December 2022. <https://www.medischcontact.nl/nieuws/laatste-nieuws/vandaag-op-de-werkvloer-1/werkvloer/mensen-hebben-het-recht-om-te-vragen-om-het-astrazeneca-vaccin.htm>.
- Hage, Jaap C., R. E. Leenes, and Arno R. Lodder. 'Hard Cases: A Procedural Approach'. *Artificial Intelligence and Law* 2, nr. 2 (1994): 113–67.
- Haibe-Kains, Benjamin, George Alexandru Adam, Ahmed Hosny, Farnoosh Khodakarami, Levi Waldron, Bo Wang, Chris McIntosh, et al 'Transparency and Reproducibility in Artificial Intelligence'. *Nature* 586, nr. 7829 (October 2020): E14–16.
- Hall, Kathryn T. *Placebos*. MIT press, 2022.
- 'Handreiking Samenwerking huisarts en specialist ouderengeneeskunde: Samenhangende geneeskundige zorg voor patiënten met een complexe zorgbehoefte'. LHV en VerenSo, 2020.
- Hart, H. L. A. 'Positivism and the Separation of Law and Morals'. *Harvard Law Review* 71, nr. 4 (1958): 593–629.
- Hattenstone, Simon. 'Lady Hale: "My Desert Island Judgments? Number One Would Probably Be the Prorogation Case"'. *The Guardian*, 11 January 2020, sec. Law. <https://www.theguardian.com/law/2020/jan/11/lady-hale-desert-island-judgments-prorogation-case-simon-hattenstone>.
- Heald, David. 'Variations of Transparency'. In *Transparency: The Key to Better Governance?*, edited by Christopher Hood en David Heald. Oxford, UK: Oxford University Press, 2011.
- Héman, René. 'Niet stiekem'. *KNMG - Actualiteit en Opinie* (blog), March 2018.
- Hendriks, Aart. 'Challenges and Obstacles to Access to Justice in Health Care'. *Recht Der Werkelijkheid* 36, nr. 3 (November 2015): 127–38.
- . 'In Beginsel (2005)'. In *Oratiebundel Gezondheidsrecht: verzamelde redes 1971-2011*. Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012.
- Hern, Alex. 'TechScape: This Cutting Edge AI Creates Art on Demand – Why Is It so Contentious?' *The Guardian*, 4 May 2022, sec. Technology. <https://www.theguardian.com/technology/2022/may/04/techscape-openai-dall-e-2>.

- Hielkema, David. 'Gediscrimineerde medewerkers en eenzijdige blik bij diagnoses: Amsterdam gaat racisme in de zorg aanpakken'. *Het Parool*, 28 February 2022, sec. Amsterdam. <https://www.parool.nl/gs-bb761baf>.
- Hilbrink, Coen. *'In het belang van het Nederlandse volk...': over de medewerking van de ambtelijke wereld aan de Duitse bezettingspolitiek 1940-1945*. 's-Gravenhage: Sdu Uitgeverij Koninginnegracht, 1995.
- Hildebrandt, Mireille. 'Law As an Affordance: The Devil Is in the Vanishing Point(s)'. *Critical Analysis of Law* 4, nr. 1 (2017).
- . 'Learning as a Machine. Crossovers Between Humans and Machines'. *Journal of Learning Analytics* 4, nr. 1 (2017): 6–23.
- Hirsch Ballin, Ernst. *Advanced Introduction to Legal Research Methods*. Edward Elgar Publishing, 2020.
- 'Hoe hoort het eigenlijk? Passend contact tussen overheid en burger'. Raad voor het openbaar bestuur, June 2014.
- Hoeren, Thomas, and Maurice Niehoff. 'Artificial Intelligence in Medical Diagnoses and the Right to Explanation'. *European Data Protection Law Review* 4, nr. 3 (2018): 308–19.
- Hoffmann, Anna Lauren. 'Where fairness fails: data, algorithms, and the limits of antidiscrimination discourse'. *Information, Communication & Society* 22, nr. 7 (7 June 2019): 900–915.
- Holroyd, Jules. 'Responsibility for Implicit Bias'. *Journal of Social Philosophy* 43, nr. 3 (2012): 274–306.
- Huisartsopleiding Nederland. *Competentieprofiel van de huisarts*, 2016. https://www.huisartsopleiding.nl/images/opleiding/Competentieprofiel_van_de_huisarts_2016.pdf.
- . *Thema's en KBA's*, 2016.
- Humphreys, Rachel, Nazia Parveen, and Annabel Sowemimo. 'Vaccine Hesitancy: What Is behind the Fears Circulating in BAME Communities?' *The Guardian - Today in Focus*, 26 January 2021, <https://www.theguardian.com/news/audio/2021/jan/26/vaccine-hesitancy-what-is-behind-the-fears-circulating-in-bame-communities-podcast>.
- Hupe, Peter, en Aurélien Buffat. 'A Public Service Gap: Capturing contexts in a comparative approach of street-level bureaucracy'. *Public Management Review* 16, nr. 4 (May 2014): 548–69.
- Hustvedt, Siri. *A Woman Looking at Men Looking at Women: Essays on Art, Sex, and the Mind*. New York: Simon & Schuster, 2016.
- . 'The Delusions of Certainty'. In *A Woman Looking at Men Looking at Women: Essays on Art, Sex, and the Mind*. New York: Simon & Schuster, 2016.
- Hutchinson, Terry, en Nigel Duncan. 'Defining and Describing What We Do: Doctrinal Legal Research'. *Deakin Law Review* 17, nr. 1 (1 October 2012): 83–119.
- Huutoniemi, Katri. 'Interdisciplinarity as Academic Accountability: Prospects for Quality Control Across Disciplinary Boundaries'. *Social Epistemology* 30, nr. 2 (2016): 163–85.
- MEL Magazine. "'I'm Overdue for a Discussion About My Role in Inspiring 'Edgelord' Shit": A Conversation with Steve Albini', 8 November 2021. <https://melmagazine.com/en-us/story/steve-albini-counsel-culture-interview>.

- ‘Incomprehensible Government’, Summary of the 2012 Annual Report of the National Ombudsman of The Netherlands’. Nationale Ombudsman, 2012.
- Verdieping - College voor de Rechten van de Mens. ‘Interview: “Laten we nou vooral leren van gemaakte fouten en kijken of we algoritmes wél verantwoord kunnen inzetten” -’. Ministerie van Algemene Zaken, 7 July 2022. <https://www.mensenrechten.nl/actueel/toegelicht/interviews/2022/laten-we-nou-vooral-leren-van-gemaakte-fouten-en-kijken-of-we-algoritmes-wel-verantwoord-kunnen-inzetten>.
- Irani, Lilly, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E. Grinter. ‘Postcolonial Computing: A Lens on Design and Development’. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1311–20. CHI ’10. New York, NY, USA: ACM, 2010. <https://doi.org/10.1145/1753326.1753522>.
- ‘Jaarverslag tuchtklachtfunctionarissen 2019’. Rapport. Ministry of Public Health, Well-Being and Sports, 30 June 2020.
- ‘Jaarverslag van de Rechtspraak’. Raad voor de Rechtspraak, May 2022.
- Jansen, Sabine. ‘Trots of Schaamte? Het vervolg’. COC Nederland, March 2022.
- Janssen, N B A T, F J Oort, P Fockens, D L Willems, H C J M de Haes, and E M A Smets. ‘Under What Conditions Do Patients Want to Be Informed about Their Risk of a Complication? A Vignette Study’. *Journal of Medical Ethics* 35, nr. 5 (1 May 2009): 276–82.
- Jensma, Folkert. ‘De Raad van State gaat nat, maar is het genoeg?’ *NRC*, 27 November 2021. <https://www.nrc.nl/nieuws/2021/11/27/zelfreflectie-bij-de-raad-van-state-over-toeslagen-hakt-erin-a4066993>.
- Jones, Meg Leta. ‘The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood’. *Social Studies of Science* 47, nr. 2 (1 April 2017): 216–39.
- Jongepier, Fleur. ‘Opinie: Jongeren voorrang geven op ic? Het draaiboek van “code zwart” rammelt aan alle kanten’. *de Volkskrant*, 17 June 2020, sec. Opinie. <https://www.volkskrant.nl/gs-b98d9cd4>.
- ‘Kabinet wil geen leeftijdselectie op intensive care, artsen houden vast aan draaiboek’, 5 January 2021. <https://nos.nl/artikel/2363133-kabinet-wil-geen-leeftijdselectie-op-intensive-care-artsen-houden-vast-aan-draaiboek>.
- Kaminski, Margot E. ‘The Right to Explanation, Explained’, 15 June 2018.
- Kaminski, Margot E., en Gianclaudio Malgieri. ‘Algorithmic impact assessments under the GDPR: producing multi-layered explanations’. *International Data Privacy Law* 11, nr. 2 (2021).
- Kammer, Claudia, en Liza van Lonkhuyzen. ‘Oud-minister Bussemaker gelooft niet meer in de participatiemaatschappij’. *NRC Handelsblad*, 14 February 2019. <https://www.nrc.nl/nieuws/2019/02/14/misschien-was-ik-naief-a3654165>.
- Kastelein, W. R. ‘Patiëntenwetgeving: Bureaucratie of Bescherming? (2001)’. In *Oratiebundel Gezondheidsrecht: verzamelde redes 1971-2011*. Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012.
- Katz, Jay. *The Silent World of Doctor and Patient*. 1984. Johns Hopkins edition. Johns Hopkins University Press, 2002.
- Kemper, Jakko, and Daan Kolkman. ‘Transparent to whom? No algorithmic accountability without a critical audience’. *Information, Communication & Society* 22, nr. 4 (2019)

- Kidd, Ian James, José Medina, and Gaile Pohlhaus, Jr. 'Introduction to The Routledge Handbook on Epistemic Injustice'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- , red. *The Routledge Handbook of Epistemic Injustice*. Paperback edition. Routledge, 2017.
- Kleingeld, Pauline. 'On Dealing with Kant's Sexism and Racism' 2, nr. 2 (2019): 21.
- Kleinherenbrink, Annelies. 'The Politics of Plasticity: Sex and Gender in the 21st Century Brain'. Thesis, University of Amsterdam, 2016.
- Koenraad, Rens. 'Op zoek naar algemene beginselen van behoorlijk Burgerschap in het Nederlands bestuursrecht'. In *25 jaar Awb: in eenheid en verscheidenheid*, edited by A.T Marseille, Tom Barkhuysen, Willemien den Ouden, Hans Peters, en Raymond Schlössels. Deventer: Wolters Kluwer, 2019.
- Kooiman, Stephanie A. 'Realtime-inzage via het patiëntenportaal'. *Nederlands Tijdschrift voor Geneeskunde*, 4.
- Koops, B.J. 'On decision transparency, or how to enhance data protection after the computational turn.' In *Privacy, due process and the computational turn*, edited by M Hildebrandt en K de Vries, 169–220. Abindon: Routledge, 2013.
- Kotzee, Ben. 'Education and Epistemic Injustice'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Kranenburg, Janny. "'The facts: Administrative versus Civil Law Courts,'" In *Motiveren, Over het motiveren van rechterlijke uitspraken*, edited by Marion Beckers, Toine De Bie, Dolf Van Harinxma thoe Slooten, René Klomp, en Anneke Rang. Prinsengrachtreeks. Ars Aequi Libri, 2017.
- Kruiter, Harry. 'De Algemene Beginselen van Behoorlijk Maatwerk'. *Instituut Publieke Waarden* (blog), 1 December 2016. <https://publiekewaarden.nl/de-algemene-beginselen-van-behoorlijk-maatwerk/>.
- Kruyswijk, Marc. 'UvA stopt met verkoop concentratiepillen na klacht'. *Het Parool*, 6 September 2018, sec. Voorpagina. <https://www.parool.nl/gs-bfaeef85>.
- Kulk, Stefan, en Stijn Van Deursen. 'Juridische aspecten van algoritmen die besluiten nemen. Een verkennend onderzoek'. Den Haag: WODC, 2020. <https://www.uu.nl/sites/default/files/tk-bijlage-onderzoeksrapport-juridische-aspecten-van-algoritmen-die-besluiten-nemen.pdf>.
- Laan, Anna Laura van der, en Gert Olthuis. 'Speuren, puzzelen of afstemmen. Alzheimerdiagnostiek'. In *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen. Boom Filosofie, 2013.
- Lagewaard, T. J. 'Epistemic Injustice and Deepened Disagreement'. *Philosophical Studies* 178, nr. 5 (1 May 2021): 1571–92.
- Lammers, Esther. 'Oud-ombudsman Alex Brenninkmeijer ziet in de toeslagenaffaire geen bedrijfsongeluk maar een falend systeem'. *Trouw*, 31 December 2020, sec. politiek. <https://www.trouw.nl/gs-bdc55fe5>.
- Landesverwaltungsgericht Wien, VGW-101/042/791/2020-44 (Landesverwaltungsgericht Wien 11 February 2022).

- Lanzing, Marjolein. “‘Strongly Recommended’ Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies”. *Philosophy & Technology*, 6 June 2018, 1–20.
- Larus, James, Chris Hankin, Siri Granum Carson, Markus Christen, Silvia Crafa, Oliver Grau, Claude Kirchner, et al ‘When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making’. White paper. Informatics Europe & EUACM, 1 January 2018. <https://doi.org/10.1145/3185595>.
- Leenen, H.H.J., J.K.M. Gevers, J. Legemaate, M.C. Ploem, M.F. Van der Mersch, E. Plomp, V.E.T. Dörenberg, en E.J.C De Jong. *Handboek gezondheidsrecht*. Edited by J Legemaate. 8e edition. Den Haag: Boom Uitgevers, 2020.
- Legemaate, J. ‘Aanpassingen van de WGBO’. *Tijdschrift voor Gezondheidsrecht* 42, nr. 6 (December 2018): 556–64.
- . ‘De informatierechten van de patiënt: te weinig en te veel’. *Tijdschrift voor Gezondheidsrecht* 35, nr. 6 (2011): 478–86.
- Legemaate, J. *Goed recht: de betekenis en de gevolgen van het recht voor de praktijk van de hulpverlening*. Preadvies Vereniging voor Gezondheidsrecht. Utrecht, 1994.
- Legemaate, J. ‘Rechtstekorten in het gezondheidsrecht’. *Tijdschrift voor Gezondheidsrecht* 42, nr. 3 (2018): 193–203.
- Legemaate, Johan. ‘Nieuwe Verhoudingen in de Spreekkamer: Juridische aspecten’. Achtergrondstudie RVZ-advies, 7 February 2013.
- . ‘Patiëntveiligheid en patiëntenrechten (2006)’. In *Oratiebundel Gezondheidsrecht: verzamelde rede 1971-2011*. Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012.
- . ‘The Development and Implementation of Patients’ Rights: Dutch Experience of the Right to Information Patients’ Rights’. *Medicine and Law* 21, nr. 4 (2002): 723–34.
- . ‘Triage bij “code zwart” in zorg kan niet anoniem’. NRC. Consulted 28 April 2021. <https://www.nrc.nl/nieuws/2021/04/26/triage-bij-code-zwart-in-zorg-kan-niet-anoniem-a4041319>.
- Legemaate, Johan, en Guy Widdershoven, red. *Basisboek ethiek en recht in de gezondheidszorg*. Boom Filosofie, 2016.
- Legemate, Dink A., en Johan Legemaate. ‘Het preoperatief informed consent’. *Nederlands Tijdschrift voor Geneeskunde* 2010, nr. 154:A2492 (2010).
- Lepore, Jill. *If Then: How the Simulmatics Corporation Invented the Future*. Illustrated edition. New York: Liveright, 2020.
- ‘Lessen uit de kinderopvangtoeslagzaken. Reflectierapport van de Afdeling Bestuursrechtspraak van de Raad van State’. Overzichtspagina. Afdeling Bestuursrechtspraak van de Raad van State,
- Levy, Karen, en Solon Barocas. ‘Designing Against Discrimination in Online Markets’. *Berkeley Technology Law Journal* 32 (2017): 1183–1238.
- Liebow, Nabina. ‘Internalized Oppression and Its Varied Moral Harms: Self-Perceptions of Reduced Agency and Criminality’. *Hypatia* 31, nr. 4 (2016): 713–29.
- Lipsky, Michael. *Street Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russel Sage Foundation, 2010.

- Loughlin, Michael. 'Epistemology, Biology and Mysticism: Comments on "Polanyi's Tacit Knowledge and the Relevance of Epistemology to Clinical Medicine"'. *Journal of Evaluation in Clinical Practice* 16, nr. 2 (2010): 298–300.
- Lukas van den Berge. 'Bestuursrecht Tussen Autonomie en Verhouding: Naar een Relationeel Bestuursrecht'. Utrecht University, 2016.
- Macarthur R. S. and Elley W. B. 'The reduction of socioeconomic bias in intelligence testing'. *British Journal of Educational Psychology* 33, nr. 2 (13 May 2011): 107–19.
- Mackor, Anne Ruth. 'Rechtsregels en medische richtlijnen. Een rechtsfilosofisch perspectief op de aard en functie van regels'. In *Medische Aansprakelijkheid*, edited by S. Heirman, E.C. Huijsmans, en R. Van den Munckhof. Kenniscentrum Milieu en Gezondheid. Wolf Legal Publishers, 2016.
- Male, R.M. van. 'Bestuursrechtspraak bij erosie van het legaliteitsbeginsel'. *Nederlands Tijdschrift voor Bestuursrecht* 2019, nr. 2 .
- . 'Van motiveringscontrole naar bestuursrechtelijke rechtsvinding'. *Nederlands Tijdschrift voor Bestuursrecht* 2007 .
- Malgieri, Gianclaudio. 'Automated Decision-Making in the EU Member States Laws: The Right to Explanation and Other "Suitable Safeguards"'. *Computer Law & Security Review* 35, nr. 5 (forthcoming , Available at SSRN: <http://dx.doi.org/.2139/ssrn.3233611> 2019).
- Malgieri, Gianclaudio, en Giovanni Comandé. 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation'. *International Data Privacy Law* 7, nr. 3 (13 November 2017).
- Malgieri, Gianclaudio, and Frank A. Pasquale. 'From Transparency to Justification: Toward Ex Ante Accountability for AI'. SSRN Scholarly Paper. Rochester, NY, 3 May 2022. <https://doi.org/10.2139/ssrn.4099657>.
- Malik, Momin M. 'A Hierarchy of Limitations in Machine Learning'. *ArXiv Preprint ArXiv:2002.05193.*, 29 February 2020. <http://arxiv.org/abs/2002.05193>.
- Marcum, James A. 'Clinical Decision-Making, Gender Bias, Virtue Epistemology, and Quality Healthcare'. *Topoi* 36, nr. 3 (1 September 2017): 501–8.
- Marseille, A.T, Tom Barkhuysen, Willemien den Ouden, Hans Peters, en Raymond Schlössels, red. *25 jaar Awb: in eenheid en verscheidenheid*. Deventer: Wolters Kluwer, 2019.
- Marseille, A.T., B.W.N. De Waard, A. Tollenaar, P. Laskewitz, en C. Boxum. 'De praktijk van de Nieuwe zaaksbehandeling in het bestuursrecht',
- Marseille, A.T., en M.F. Vermaat. 'Burgers op zoek naar rechtsbescherming in het sociaal domein'. *Handicap & Recht* 1, nr. 1 (June 2017): 9–15.
- Marseille, Bert, en Alex Brenninkmeijer. 'Een dialoog met de Raad van State na de toelagenaffaire'. *NJB* 2021, nr. 8 (26 February 2021): 575.
- Mattu, Jeff Larson, Julia Angwin, Lauren Kirchner, Surya. 'How We Analyzed the COMPAS Recidivism Algorithm'. ProPublica. Consulted 15 December 2020. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=XSO7CCiM7D0udJrFYQeZnvAitR3ZT0sj>.
- Matute, Helena, Fernando Blanco, Ion Yarritu, Marcos Díaz-Lago, Miguel A. Vadillo, en Itxaso Barberia. 'Illusions of causality: how they bias our everyday thinking and how they could be reduced'. *Frontiers in Psychology* 6 (2 July 2015): 888.

- McClure, Tess. 'New Zealand Passes Plain Language Bill to Jettison Jargon'. *The Guardian*, 19 October 2022, sec. World news. <https://www.theguardian.com/world/2022/oct/20/new-zealand-passes-plain-language-bill-to-jettison-jargon>.
- McCrudden, C. 'Human Dignity and Judicial Interpretation of Human Rights'. *European Journal of International Law* 19, nr. 4 (1 September 2008): 655–724.
- McQuillan, Dan. *Resisting AI: An Anti-Fascist Approach to Artificial Intelligence*. Bristol University Press, 2022.
- Medina, José. 'Varieties of Hermeneutical Injustice'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, en Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Mein, A.G., en Bert Marseille. 'Informeel aanpak bij bezwaar: rapportage werkpakket 2: de belfase'. Amsterdam University of Applied Sciences, AKMI, 2019.
- Metcalfe, Jacob, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, en Madeleine Clare Elish. 'Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts'. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 735–46. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. <https://doi.org/10.1145/3442188.3445935>.
- Meuwese, A.C.M. 'Grip op normstelling in het datatijdperk'. In *Algemene regels in het bestuursrecht*. Preadviezen Vereniging voor Bestuursrecht, VAR-reeks 158. Boom Juridisch, 2017.
- Miller, Tim. 'Explanation in Artificial Intelligence: Insights from the Social Sciences', 22 June 2017. <https://arxiv.org/abs/1706.07269>.
- Miller, Tim, Piers Howe, and Liz Sonenberg. 'Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences', nr. arXiv:1712.00547 [cs.AI] (2 December 2017). <https://arxiv.org/abs/1712.00547>.
- Mills, Charles W. "'Ideal Theory" as Ideology'. *Hypatia* 20, nr. 3 (2005): 165–84.
- . 'White Ignorance'. In *Agnology: The Making and Unmaking of Ignorance*, edited by Robert N. Proctor en Londa Schiebinger. Stanford University Press, 2008.
- Ministerie van Algemene Zaken. 'Memorie van Toelichting wetsvoorstel Wet versterking waarborgfunctie Awb (pre-consultatieversie 18 January 2023)'. Ministerie van Algemene Zaken, 18 January 2023.
- Ministerie van Binnenlandse Zaken. 'Besluit van 3 May 2022, houdende instelling van een staatscommissie tegen discriminatie en racisme'. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties,
- . 'Kamerbrief met Kabinetsreactie "Hoe hoort het eigenlijk? Passend contact tussen overheid en burger"', 9 March 2016.
- Ministerie van Financiën, Ministerie van. 'Kamerbrief over Fraudesignaleringsvoorziening en vraagstuk institutioneel racisme'. Ministerie van Algemene Zaken, 30 May 2022. <https://www.rijksoverheid.nl/documenten/kamerstukken/2022/05/30/kamerbrief-reactie-op-verzoeken-over-fraudesignaleringsvoorziening>.
- Ministerie van Justitie en Veiligheid. 'Antwoorden Kamervragen over discriminatie en racisme bij de politie', 6 July 2022.

- Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, en Timnit Gebru. 'Model Cards for Model Reporting'. *ArXiv:1810.03993 [Cs]*, 5 October 2018. <http://arxiv.org/abs/1810.03993>.
- Mitova, Veli. 'A New Argument for the Non-Instrumental Value of Truth'. *Erkenntnis* 2021 (12 June 2021): 1–23.
- . 'Decolonising Knowledge Here and Now'. *Philosophical Papers* 49, nr. 2 (3 May 2020): 191–212.
- . 'Explanatory Injustice and Epistemic Agency'. *Ethical Theory and Moral Practice* 23, nr. 5 (November 2020): 707–22.
- Moerel, Lokke, and Marijn Storm. 'Automated Decisions Based on Profiling - Information, Explanation or Justification? That Is the Question!' *Oxford Law Faculty, Law and Autonomous Systems Series* (blog), 27 April 2018. <https://www.law.ox.ac.uk/business-law-blog/blog/2018/04/law-and-autonomous-systems-series-automated-decisions-based-profiling>.
- Mohamed, Shakir, Marie-Therese Png, and William Isaac. 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence'. *Philosophy & Technology*, 12 July 2020.
- Mol, Annemarie. *The Body Multiple: Ontology in Medical Practice*. Duke University Press, 2003.
- . *The Logic of Care: Health and the Problem of Patient Choice*. Routledge, 2008.
- Mol, Annemarie, en Peter van Lieshout. *Ziek is het woord niet: medicalisering, normalisering en de veranderende taal van huisartsengeneeskunde en geestelijke gezondheidszorg, 1945-1985*, 2008.
- Montague, Jules. 'What Happens When Doctors Change Your Diagnosis?' *The Guardian*, 11 June 2018, sec. Life and style. <http://www.theguardian.com/lifeandstyle/2018/jun/11/what-happens-when-doctors-change-your-diagnosis>.
- Montesquie Instituut / kenniscentrum parlementaire democratie. 'Trias politica: machtscheiding en machtspreiding'. Consulted 11 November 2022. https://www.montesquieu-instituut.nl/id/vhnmmt7lidzx/trias_politica_machtscheiding_en.
- Moss, Emanuel, Elizabeth Anne Watkins, Ranjit Singh, Madeleine Clare Elish, en Jacob Metcalf. 'Assembling Accountability: Algorithmic Impact Assessment for the Public Interest'. *Data & Society*, 29 June 2021.
- 'Motie van de leden Bouchallikh en Dekker-Abudlaziz, KamerstukkenTweede Kamer, vergaderjaar 2021–2022, 26 643, nr. 835', March 2022.
- Moyn, Samuel. *Not Enough: Human Rights in an Unequal World*. Cambridge, Massachusetts: Harvard University Press, 2019.
- Muhammad, Selma. 'The Fairness Handbook', 17 May 2022. <http://amsterdamintelligence.com/resources/the-fairness-handbook>.
- Muris, Jean, Roger Damoiseaux, en Nynke van Dijk. 'Bijwerkingen en valkuilen van EBM: trap er niet in!' *Huisarts en wetenschap* 60, nr. 11 (November 2017): 548–51.
- Narayan, Nitin 'A Decision Support System for the Court of East Brabant'. Professional Doctorate in Engineering, Jheronimus Academy of Data Science, 2019.

- Neal, Mary. 'Respect for Human Dignity as "Substantive Basic Norm"'. *International Journal of Law in Context* 10, nr. 1 (March 2014): 26–46.
- Neyland, Daniel. 'Bearing Account-Able Witness to the Ethical Algorithmic System'. *Science, Technology, & Human Values* 41, nr. 1 (January 2016): 50–76.
- Nisar, Muhammad Azfar, and Ayesha Masood. 'Bureaucracy and the Other: A Systematic Review of Postcolonial Scholarship in Public Administration'. SSRN Scholarly Paper. Rochester, NY, 14 July 2021. <https://doi.org/10.2139/ssrn.3886409>.
- Nissenbaum, Helen. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2009.
- NJCM and others v. The Netherlands (English) ECLI:NL:RBDHA:2020:1878 (The Hague District Court 6 March 2020).
- 'No evidence without context About the illusion of evidence-based practice in healthcare'. Publicatie. Raad voor Volksgezondheid en Samenleving, 2017.
- Noorman, F.M. 'Quality and Administration of the Dutch Social Security System: An Impression'. In *Quality of Decision-Making in Public Law: Studies in Administrative Decision-Making in the Netherlands*, edited by K. J. de Graaf, J H Jans, A T Marseille, and J de Ridder, Europa Law Publishing 2007
- Nordell, Jessica. 'The Bias That Blinds: Why Some People Get Dangerously Different Medical Care'. *The Guardian*, 21 September 2021, sec. Science. <https://www.theguardian.com/science/2021/sep/21/bias-that-blinds-medical-research-treatment-race-gender-dangerous-disparity>.
- NOS.nl. 'Yesilgöz botst fel met Simons over nasleep Sint-intocht Staphorst', 22 November 2022. <https://nos.nl/artikel/2453437-yesilgoz-botst-fel-met-simons-over-nasleep-sint-intocht-staphorst>.
- Nyrup, Rune, and Diana Robinson. 'Explanatory Pragmatism: A Context-Sensitive Framework for Explainable Medical AI'. *Ethics and Information Technology* 24, nr. 1 (2022): 13.
- Oerlemans, J., en Y.E. Schuurmans. 'Internetonderzoek door bestuursorganen'. *Nederlands Juristenblad* 2019, nr. 20 (2019).
- 'Onderzoek effecten FSV Toeslagen'. Rapport. Price Waterhouse Coopers, November 2021.
- O'Neill, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2016.
- O'Neill, O. 'Some Limits of Informed Consent'. *Journal of Medical Ethics* 29, nr. 1 (1 February 2003).
- O'Neill, Onora. 'Accountability, Trust and Informed Consent in Medical Practice and Research'. *Clinical Medicine* 4, nr. 3 (1 May 2004): 269–76.
- 'Ongeskend Onrecht: Verslag van de Parlementaire ondervragingscommissie Kinderopvangtoeslag'. Den Haag: Tweede Kamer der Staten-Generaal, 17 December 2020.
- 'Onggevraagd advies over de effecten van de digitalisering voor de rechtsstatelijke verhoudingen'. Raad van State, 31 August 2018. Kamerstukken II 2017/18, 26643, nr. 557.
- Oostveen, Manon, and Kristina Irion. 'The Golden Age of Personal Data: How to Regulate an Enabling Fundamental Right?' In *Personal Data in Competition, Consumer Protection and IP Law - Towards a Holistic Approach?*, edited by Bakhoun, Conde Callego, Mackenordt, en Surblyte. Berlin: Springer, 2017.

- ‘Opnemen van gesprekken door patienten: Uitkomsten raadpleging KNMG Artsenpanel’. KNMG,
- ‘Opnemen van het gesprek’. Consulted 27 January 2021. <https://www.knmg.nl/advies-richtlijnen/dossiers/opnemen-van-het-gesprek.htm>.
- Osselen, Esther van, Ron Helsloot, Emma van Zalinge, en Ger van der Werf. *Geschiedenis van de Huisartsgeneeskunde*. Utrecht: Nederlands Huisartsen Genootschap, 2016.
- Oswald, Marion. ‘Algorithm-assisted decision-making in the public sector: framing the issues using administrative law rules governing discretionary power’. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, nr. 2128 (13 September 2018): 20170359.
- Ouden, Willemien den, en Heleen van Amerongen. ‘Het bestuursorgaan-begrip voorbij?’ In *De conclusie voorbij. Liber amicorum aangeboden aan Jaap Polak*, edited by M Bosma, B.J. van Eetkoven, O van Loon, H.G. Lubberdink, J.C.A. de Poorter, en B.J. Schueler. Ars Aequi, 2017.
- Overkleef-Verburg, G. ‘Basisregistraties en rechtsbescherming. Over de dualisering van de bestuursrechtelijke rechtsbetrekking’. *NTB* 2009, nr. 10 . Consulted 27 May 2019.
- Palmboom, G. G., D. L. Willems, N. B. a. T. Janssen, and J. C. J. M. de Haes. ‘Doctor’s Views on Disclosing or Withholding Information on Low Risks of Complication’. *Journal of Medical Ethics* 33, nr. 2 (1 February 2007): 67–70.
- Parlementaire onderzoekscommissie effectiviteit antidiscriminatiewetgeving. ‘Gelijk recht doen: Een parlementair onderzoek naar de mogelijkheden van de wetgever om discriminatie tegen te gaan’. Eerste Kamer der Staten-Generaal, Consulted 22 September 2022.
- Peeters, M.J.M. ‘Hoe wordt de discretionaire bevoegdheid in schrijnende situaties gebruikt?’ *A&MR* 2018, nr. 1 : 7.
- Pellegrino, Edmund D. ‘Praxis as a Keystone for the Philosophy and Professional Ethics of Medicine: The Need for an Arch-Support: Commentary on Toulmin and Wartofsky’. In *Philosophy of Medicine and Bioethics: A Twenty-Year Retrospective and Critical Appraisal*, edited by Ronald A. Carson and Chester R. Burns, 69–84. Philosophy and Medicine. Dordrecht: Springer Netherlands, 1997.
- Philip Alston, United Nations Special Rapporteur on extreme poverty and human rights in the case of NJCM /De Staat der Nederlanden (C/09/550982/HA ZA 18/388). ‘Brief by the United Nations Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM C.s./De Staat der Nederlanden (SyRI) before the District Court of The Hague (case number: C/09/550982/HA ZA 18/388)’, 26 September 2019. <https://www.ohchr.org/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf>.
- Pierce, Robin L. ‘Medical Privacy: Where Deontology and Consequentialism Meet’. In *The Handbook of Privacy*, edited by Bart van der Sloot and Aviva de Groot. Amsterdam University Press, 2018.
- Pierce, Robin, Sigrid Sterckx, and Wim Van Biesen. ‘A Riddle, Wrapped in a Mystery, inside an Enigma: How Semantic Black Boxes and Opaque Artificial Intelligence Confuse Medical Decision-Making’. *Bioethics* 36, nr. 2 (2022): 113–20.

- Pim. 'Protocol: geef zorgverleners en jongeren voorrang bij extreme druk op de IC - NRC'. *NRC*, 16 June 2020. <https://www.nrc.nl/nieuws/2020/06/16/protocol-geef-zorgverleners-en-jongeren-voorrang-bij-extreme-druk-op-de-ic-a4002978>.
- 'Plenair verslag Tweede Kamer, 40e vergadering', 5 January 2021. https://www.tweedekamer.nl/kamerstukken/plenaire_verslagen/detail/2020-2021/40.
- Pohlhaus, Jr., Gaile. 'Varieties of Epistemic Injustice'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Powles, Julia, and Hal Hodson. 'Google DeepMind and Healthcare in an Age of Algorithms'. *Health and Technology* 7, nr. 4 (1 December 2017): 351–67.
- Prince, Russell. 'The Geography of Statistics: Social Statistics from Moral Science to Big Data'. *Progress in Human Geography* 44, nr. 6 (15 September 2019).
- Prins, Corien, Dennis Broeders, Henk Griffioen, Anne-Greet Keizer, and Esther Keymolen. 'IGovernment - Synthesis of WRR Report 86 (English Version)'. The Netherlands Scientific Council for Government Policy, 15 March 2011.
- . 'IGovernment (English Version)'. The Netherlands Scientific Council for Government Policy, 15 March 2011.
- Proctor, Robert N. 'The Nazi Campaign against Tobacco'. In *Medicine and Medical Ethics in Nazi Germany: Origins, Practices, Legacies*, edited by Francis R. Nicosia en Jonathan Huener, 1st Edition. New York: Berghahn Books, 2002.
- Proctor, Robert N., and Londa Schiebinger, red. *Agnology: The Making and Unmaking of Ignorance*. Stanford University Press, 2008.
- Proposal for a Regulation laying down harmonised rules on artificial intelligence, Pub. L. No. COM(2021) 206 final (2021).
- Raad van State. 'Aanbeveling aan bestuursrechter: pas rechterlijke evenredigheidstoets aan (ECLI:NL:RVS:2021:1468)'. *Raad van State* (blog). Raad van State. Consulted 6 February 2022. <https://www.raadvanstate.nl/actueel/nieuws/@126011/conclusie-evenredigheidstoets/>.
- 'Raamplan Medical Training Framework'. Nederlandse Federatie van Universitair Medische Centra (NFU), May 2020. https://www.nfu.nl/img/pdf/20.1577_Raamplan_Medical_Training_Framework_2020_-_May_2020.pdf.
- Rademakers, Jany en Nederlands instituut voor onderzoek van de gezondheidszorg (Utrecht). *Kennisynthese: gezondheidsvaardigheden: niet voor iedereen vanzelfsprekend*. Utrecht: NIVEL, 2014.
- Ranchordás, Sofia. 'Empathy in the Digital Administrative State'. *University of Groningen Faculty of Law Research Paper* 2021, nr. 13 (2021).
- Raso, Jennifer. 'Unity in the Eye of the Beholder? Reasons for Decision in Theory and Practice'. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 1 August 2016. <https://papers.ssrn.com/abstract=2840488>.
- 'Recht vinden bij de rechtbank: lessen uit kinderopvangtoeslagzaken'. Werkgroep reflectie toeslagenaffaire rechtbanken, October 2021.
- AD.nl. 'Rechters laken gijzeling wanbetaler door justitie', 24 February 2014, sec. Binnenland. <https://www.ad.nl/binnenland/rechters-laken-gijzeling-wanbetaler-door-justitie~aba57f1a/>.

- NRC. “Rechters zetten onze deskundigheid weg als een mening”, 17 October. <https://www.nrc.nl/nieuws/2022/10/17/rechters-zetten-onze-deskundigheid-weg-als-een-mening-a4145406>.
- Redden, Joanna, Lina Dencik, and Harry Warne. ‘Datafied child welfare services: unpacking politics, economics and power’. *Policy Studies* 41, nr. 5 (2 September 2020): 507–26.
- Regulation (EU) 2017/ 745 on medical devices (2017).
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data., Pub. L. No. 32016R0679, 119 OJ L (2016).
- Reisman, Dillon, Jason Schultz, Kate Crawford, and Meredith Whittaker. ‘Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability’. AI Now Institute, April 2018.
- Reuters. ‘Google Told Scientists to Use “a Positive Tone” in AI Research, Documents Show’. *The Guardian*, 23 December 2020, sec. Technology. <https://www.theguardian.com/technology/2020/dec/23/google-scientists-research-ai-postive-tone>.
- Richardson, Rashida. ‘Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities’. *Berkeley Technology Law Journal* 36, nr. 3 (2021).
- Roessler, Beate. *Autonomy: An Essay on the Life well-lived*. Suhrkamp Insel Bücher, last consulted on 9 December 2020.
- Romet v. the Netherlands, No. 7094/06 (ECtHR 14 February 2012).
- Romijn, Peter. *Burgemeesters in oorlogstijd: besturen tijdens de Duitse bezetting*. Amsterdam: Balans, 2006.
- Rood, Jurriën. *Lentz. De man achter het Persoonsbewijs*, 2022.
- Rosenberg, Charles E. ‘The Tyranny of Diagnosis: Specific Entities and Individual Experience’. *The Milbank Quarterly* 80, nr. 2 (June 2002): 237–60.
- Rudin, Cynthia. ‘Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead’. *Nature Machine Intelligence* 1, nr. 5 (May 2019): 206–15.
- Rule, Peter, en Vaughn Mitchell John. ‘A Necessary Dialogue: Theory in Case Study Research’. *International Journal of Qualitative Methods* 14, nr. 4 (20 November 2015): 1609406915611575.
- Sample, Ian. ‘Maths and Tech Specialists Need Hippocratic Oath, Says Academic’. *The Guardian*, 16 August 2019, sec. Science. <https://www.theguardian.com/science/2019/aug/16/mathematicians-need-doctor-style-hippocratic-oath-says-academic-hannah-fry>.
- Sax, Marijn. ‘Optimization of What? For-Profit Health Apps as Manipulative Digital Environments’. *Ethics and Information Technology*, 3 January 2021.
- Saxena, Devansh, Charles Repaci, Melanie D Sage, and Shion Guha. ‘How to Train a (Bad) Algorithmic Caseworker: A Quantitative Deconstruction of Risk Assessments in Child Welfare’. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–7. New Orleans LA USA: ACM, 2022.
- Schauer, Frederick. ‘Giving Reasons’. *Stanford Law Review* 47, nr. 4 (April 1995): 633–59.

- Scheltema, M. 'Een wet van Meden en Perzen? Geen onwrikbare 15 wet in het hedendaags bestuursrecht'. In: *Wetgeving en uitvoering Nederlandse Vereniging voor Wetgeving*. Nederlandse Vereniging voor Wetgeving, 2021.
- . 'Wetgeving in de responsieve rechtsstaat'. *RegelMaat* 33, nr. 3 (May 2018): 120–31.
- Scheltema, Tatjana. 'Wordt een advocaat slechts een privilege voor de rijken?' *De Groene Amsterdammer*, 1 February 2019. <https://www.groene.nl/artikel/een-leemte-in-de-rechtshulp>.
- Schermer, Maartje. 'Wat is Ziek, Wat is Gezond?' In *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen. Boom Filosofie, 2013.
- Schlössels, Raymond. 'De Harde Kern van Behoorlijkheid: Over rechtmatigheid, behoorlijkheid en de Nationale ombudsman', Nijmegen Migration Law Working Papers Series, 2014, nr. 3 (2014).
- Schlössels, R.J.N. 'Discretionaire dogmatiek... anders de Afdeling bestuursrechtspraak?' *Nederlands Tijdschrift voor Bestuursrecht* 2018, nr. 52 (2018).
- . 'Kroniek beginselen van behoorlijk bestuur 2000'. *Nederlands Tijdschrift voor Bestuursrecht* 2000, nr. 9 (2000).
- . 'Kroniek beginselen van behoorlijk bestuur 2006'. *Nederlands Tijdschrift voor Bestuursrecht* 2006, nr. 6 (2006).
- . 'Kroniek beginselen van behoorlijk bestuur 2008'. *Nederlands Tijdschrift voor Bestuursrecht* 2008, nr. 3 (2008).
- . 'Kroniek beginselen van behoorlijk bestuur 2010'. *Nederlands Tijdschrift voor Bestuursrecht* 2010, nr. 7 (2010).
- . 'Kroniek beginselen van behoorlijk bestuur 2013'. *Nederlands Tijdschrift voor Bestuursrecht* 2013, nr. 10 (2013).
- . 'Kroniek beginselen van behoorlijk bestuur 2016'. *Nederlands Tijdschrift voor Bestuursrecht* 2016 (2016).
- . 'Kroniek beginselen van behoorlijk bestuur 2017'. *Nederlands Tijdschrift voor Bestuursrecht* 2017, nr. 9 (2017).
- . 'Kroniek beginselen van behoorlijk bestuur 2018'. *Nederlands Tijdschrift voor Bestuursrecht* 2018, nr. 9 (21 October 2018).
- Schlössels, R.J.N., en S.E. Zijlstra. *Bestuursrecht in de sociale rechtsstaat 1*. 6e druk. Wolters Kluwer, 2010.
- Schuurmans, Y.E. 'Bewijslastverdeling in het bestuursrecht: zorgvuldigheid en bewijsvoering beschikkingen'. Vrije Universiteit van Amsterdam, 2006.
- . 'Toeslagenaffaire: outlier of symptoom van het systeem?' *Rechtsgeleerd Magazijn Themis* 2021, nr. 6 : 205–7.
- Selbst, Andrew D., en Solon Barocas. 'The Intuitive Appeal of Explainable Machines'. 87 *Fordham Law Review* 1085, 2018.
- Selbst, Andrew D., and Julia Powles. 'Meaningful Information and the Right to Explanation'. *International Data Privacy Law* 7, nr. 4 (1 November 2017): 233–42.
- Sevil, Malika, en Jop van Kempen. 'Amsterdamse huisartsen prikken soms door: "Anders is het mensonterend"'. *Het Parool*, 14 April 2021, sec. Amsterdam. <https://www.parool.nl/gsb2c8a55a>.

- Medisch Contact. 'Shared decision making is drijfzand', 9 November 2016.
- Sharon, Tamar. 'Self-tracking en sociale netwerken in de gezondheidszorg. Verschuivende definities van gezondheid en patiënt-zijn'. In *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen. Boom Filosofie, 2013.
- . 'When Digital Health Meets Digital Capitalism, How Many Common Goods Are at Stake?' *Big Data & Society* 5, nr. 2 (1 July 2018):
- Sijmons, J.G. 'De stimulerende middelen van de wetgever (2007)'. In *Oratiebundel Gezondheidsrecht: verzamelde redes 1971-2011*. Den Haag: Vereniging voor Gezondheidsrecht/SDU, 2012.
- Silven, Anna V., Petra G. van Peet, Sarah N. Boers, Monique Tabak, Aviva de Groot, Djoke Hendriks, Hendrikus J. A. van Os, et al 'Clarifying Responsibility: Professional Digital Health in the Doctor-Patient Relationship, Recommendations for Physicians Based on a Multi-Stakeholder Dialogue in the Netherlands'. *BMC Health Services Research* 22, nr. 1 (2022): 129.
- Sloane, Mona, Emanuel Moss, Olaitan Awomolo, en Laura Forlano. 'Participation is not a Design Fix for Machine Learning'. arXiv, 11 August 2020. <https://doi.org/10.48550/arXiv.2007.02423>.
- Sloot, Bart van der. 'The Practical and Theoretical Problems with "Balancing": Delfi, Coty and the Redundancy of the Human Rights Framework'. *Maastricht Journal of European and Comparative Law* 23, nr. 3 (June 2016): 439–59.
- Smaling, Elmer. 'Controversial DNA Testing? Address the Ethical Issues'. *Erasmus Magazine* (blog), 14 October 2021. <https://www.erasmusmagazine.nl/en/2021/10/14/controversial-dna-testing-address-the-ethical-issues/>.
- Smith, Barbara Ellen. 'Black Lung: The Social Production of Disease'. *International Journal of Health Services* 11, nr. 3 (1 July 1981): 343–59.
- Smith, David. 'How Did Republicans Turn Critical Race Theory into a Winning Electoral Issue?' *The Guardian*, 3 November 2021, sec. US news. <https://www.theguardian.com/us-news/2021/nov/03/republicans-critical-race-theory-winning-electoral-issue>.
- Solove, Daniel J. 'Privacy Self-Management and the Consent Dilemma'. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 4 November 2012.
- Sparrow, Robert, and Joshua Hatherley. 'The promise and perils of AI in medicine' 17 (1 December 2019): 79–109. <https://doi.org/10.24112/ijccpm.171678>.
- Stamenkovikj, Natasha. 'The Truth in Times of Transitional Justice: The Council of Europe and the Former Yugoslavia'. Tilburg Universitij, 2019.
- Staten-Generaal, Tweede Kamer der. 'Memorie van toelichting Regels inzake de gemeentelijke ondersteuning op het gebied van zelfredzaamheid, participatie, beschermd wonen en opvang (Wet maatschappelijke ondersteuning 2015) (kst-33841-3)', 13 January 2014. <https://zoek.officielebekendmakingen.nl/kst-33841-3>.
- Stellinga, Marieke, en Petra De Koning. 'PvdA vindt eigen Participatiewet mislukt'. *NRC*, 11 November 2020. <https://www.nrc.nl/nieuws/2020/11/11/pvda-vindt-eigen-participatiewet-mislukt-a4019751>.
- Stigter, Bianca. *Atlas van een bezette stad*. Atlas Contact, 2019.

- Stobbe, Emiel. 'Ervaringen na een klacht over de huisarts. Veertien diepte-interviews met patiënten'. *Huisarts & Wetenschap* 2020, nr. 63 (2020).
- Sullivan, Michael. 'Epistemic Justice and the Law'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Synced. 'Yann LeCun Quits Twitter Amid Acrimonious Exchanges on AI Bias'. *Synced* (blog), 1 July 2020. <https://syncedreview.com/2020/06/30/yann-lecun-quits-twitter-amid-acrimonious-exchanges-on-ai-bias/>.
- 'SyRI legislation in breach of European Convention on Human Rights'. Consulted on 23 September 2020. <https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Rechtbanken/Rechtbank-Den-Haag/Nieuws/Paginas/SyRI-legislation-in-breach-of-European-Convention-on-Human-Rights.aspx>.
- Taebe, Behnam, Jeroen van den Hoven, and Stephanie J. Bird. 'The Importance of Ethics in Modern Universities of Technology'. *Science and Engineering Ethics* 25, nr. 6 (1 December 2019): 1625–32.
- Tai, Eric Tjong Tjin. 'Zorg, privaatrecht en publiekrecht: van ondersteuning naar handhaving, en terug'. *Recht der Werkelijkheid*, 2010, 20.
- Tamara van Ark, Minister of Health care and Sport. 'Kamerbrief 1801920-216248-PZO over Draaiboek Triage op basis van niet-medische overwegingen voor IC-opname ten tijde van fase 3 in de COVID-19 pandemie'. Ministerie van Algemene Zaken, 4 January 2021.
- Taylor, Linnet. 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally'. *Big Data & Society* 4, nr. 2 (1 December 2017).
- . 'What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally'. *Big Data & Society* 4, nr. 2 (1 December 2017):
- Ten Hooven, Marcel. 'De verzorgingsstaat is grimmig geworden'. *De Groene Amsterdammer*, 21 January, 2021. <https://www.groene.nl/artikel/de-verzorgingsstaat-is-grimmig-geworden>.
- 'Terug aan tafel, samen de klacht oplossen: Behandeling klachten over zorg, jeugdhulp en begeleiding naar werk'. Nationale Ombudsman, Maart 2017.
- The AI and Robotics group at the Tilburg Institute for Law, Technology and Society. 'Response on the draft ethical guidelines for trustworthy AI produced by the European Commission's High-Level Expert Group on Artificial Intelligence.', 31 January 2019.
- 'The DAIR Institute'. Consulted on 20 November 2022. <https://www.dair-institute.org/about>.
- The European Group on Ethics in Science and New Technologies (EGE) | EGE - Research and Innovation - European Commission. 'Statement on Artificial Intelligence, Robotics and "Autonomous" Systems'. European Commission, March 2018.
- 'The SyRI Victory: Holding Profiling Practices to Account'. Consulted on 25 September 2020. <https://digitalfreedomfund.org/the-syri-victory-holding-government-profiling-to-account/7/>.
- T.J. Poppema. 'Commentaar bij: Algemene wet bestuursrecht, Artikel 3:46 [Deugdelijke motivering]'. In *Encyclopedie Sociale Verzekeringen, Module Uitvoering sociale zekerheid en bestuursrecht*. Deventer: Kluwer, Consulted 24 April 2019.
- 'Toekomstvisie 2012-2022 | NHG'. Consulted 27 January 2021. <https://www.nhg.org/toekomstvisie>.


- Tollenaar, A. 'Empathie in het sociaal domein'. *RegelMaat* 33, nr. 3 (May 2018): 132–42.
- Tollenaar, Albertjan. 'Bestuursrechtelijke normering en "big data"'. *Nederlands Tijdschrift voor Bestuursrecht* 2017, nr. 16 (2014).
- . 'Maintaining Administrative Justice in the Dutch Regulatory Welfare State'. *University of Groningen Faculty of Law Research Paper* 2016, nr. 24.
- Tolsma, H.D., A.T Marseille, en K.J. de Graaf. 'Prettig Contact met de Overheid 5: Juridische kwaliteit van de informele aanpak beoordeeld'. Project Prettig Contact met de Overheid. Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2013.
- De Groene Amsterdammer. 'Trainen voor de test', April 2014. <https://www.groene.nl/artikel/trainen-voor-de-test>.
- Trappenburg, Margo. 'Ik en mijn medepatiënt: Juridisering in de gezondheidszorg'. *Recht der Werkelijkheid*, 2010, 12.
- 'Triage scenario non-medical considerations IC admittance in COVID-19 pandemic phase 3 (version 2.0)'. KNMG & Federation Medical Specialists, November 2020. https://www.demedischspecialist.nl/sites/default/files/Draaiboek%20Triage%20op%20basis%20van%20niet-medische%20overwegingen%20IC-opnametvfase%203_COVID19pandemie.pdf.
- Tsosie, Rebecca. 'Indigenous Peoples, Anthropology, and the Legacy of Epistemic Injustice'. In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr., Paperback edition. Routledge, 2017.
- Tuck, Eve, and K. Wayne Yang. 'Decolonization Is Not a Metaphor'. *Decolonization: Indigeneity, Education & Society* 1, nr. 1 (2012).
- Twickler, Hoogstraaten, Reuwer, Singels, Stronks, en Essink-Bot. 'Laaggeletterdheid en beperkte gezondheidsvaardigheden vragen om een antwoord in de zorg'. *Nederlands Tijdschrift voor Geneeskunde* 2009, nr. 153:A250. Consulted 24 June 2020. <https://www.ntvg.nl/artikelen/laaggeletterdheid-en-beperkte-gezondheidsvaardigheden-vragen-om-een-antwoord-de-zorg/volledig>.
- Uitvoeringswet Algemene verordening gegevensbescherming (2018). https://www.eerstekamer.nl/behandeling/20180522/publicatie_wet/document3/f=/vkoj2ezcplyz.pdf.
- Valk, Guus. 'Nationale Ombudsman: "Laat Rutte maar een club oprichten die onze rapporten leest"'. *NRC*. Consulted 19 November 2022. <https://www.nrc.nl/nieuws/2021/05/11/nationale-ombudsman-de-afrekencultuur-bestaat-nog-altijd-a4043283>.
- 'Van Wet naar Praktijk: Implementatie van de WGBO. Deel 2: Informatie en toestemming'. Utrecht: KNMG, 2004.
- Veen, Gerrit van der. 'Digitalisering in het omgevingsrecht en mogelijke invloed op de Awb: De burger tussen de ambities en doelstellingen van de Awb'. In *25 jaar Awb: in eenheid en verscheidenheid*, edited by A.T Marseille, Tom Barkhuysen, Willemien den Ouden, Hans Peters, en Raymond Schlössels. Deventer: Wolters Kluwer, 2019.
- Verweij, Marcel, en Roland Pierik. 'Het pijnlijke gesprek over ziekenhuisbedden moet juist nu gevoerd worden'. *Bij Nader Inzien* (blog), 23 March 2020. <https://bijnaderinzien.com/2020/03/23/laat-niet-aan-artsen-over-wie-een-bed-krijgt-op-de-intensive-care/>.
- 'Verwerking van persoonsgegevens in het sociaal domein: De rol van toestemming'. Autoriteit Persoonsgegevens, April 2016.

- Vetzo, Max, Janneke Gerards, en Remco Nehmelman. *Algoritmes en Grondrechten*. Montaigne reeks. Boom Juridisch, 2018.
- Voermans, W.J.M. 'Besturen met regels, volgens de regels'. In *Algemene regels in het bestuursrecht*. Preadviezen Vereniging voor Bestuursrecht, VAR-reeks 158. Boom Juridisch, 2017.
- Vredenburg, Kate. 'Freedom at Work: Understanding, Alienation, and the AI-Driven Workplace'. *Canadian Journal of Philosophy* 52, nr. 1 (9 February 2022).
- . 'The Right to Explanation'. *Journal of Political Philosophy* 30, nr. 2 (2022): 209–29.
- Waard, Boudewijn de. 'Proportionality in Dutch administrative law'. In *The Judge and the Proportionate Use of Discretion: A Comparative Administrative Law Study*. Routledge research in EU law. New York: Routledge, 2015.
- Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation'. *International Data Privacy Law*, 2017.
- Wahlberg, Lena, and Johannes Persson. 'Importing Notions in Health Law: Science and Proven Experience'. *European Journal of Health Law* 24, nr. 5 (10 November 2017): 565–90.
- Waldron, Jeremy. 'How Law Protects Dignity'. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 15 December 2011. <https://papers.ssrn.com/abstract=1973341>.
- Walker, Margaret Urban. 'Truth Telling as Reparations'. *Metaphilosophy* 41, nr. 4 (2010): 525–45.
- Wartofsky, Marx W. 'What Can the Epistemologists Learn from the Endocrinologists? Or Is the Philosophy of Medicine Based on a Mistake?' In *Philosophy of Medicine and Bioethics: A Twenty-Year Retrospective and Critical Appraisal*, edited by Ronald A. Carson en Chester R. Burns, 55–68. Philosophy and Medicine. Dordrecht: Springer Netherlands, 1997.
- Watson, Kenneth. 'Goede zorg, informed consent & shared-decision making: nieuwe basis onder goed hulpverlenerschap en medische aansprakelijkheid?' *Letsel & Schade* 2018, nr. 3.
- 'We Sense Trouble: Automated Discrimination and Mass Surveillance in Predictive Policing in the Netherlands'. Amnesty International, 2020.
- Wear, Stephen. *Informed Consent: Patient Autonomy and Physician Beneficence within Clinical Medicine*. Clinical Medical Ethics. Springer Netherlands, 1993.
- Weinberger, David. 'Don't Make Artificial Intelligence Artificially Stupid in the Name of Transparency'. *Wired*, <https://www.wired.com/story/dont-make-ai-artificially-stupid-in-the-name-of-transparency/>.
- 'Weten is nog geen doen. Een realistisch perspectief op redzaamheid'. Wetenschappelijke Raad voor het Regeringsbeleid, 2017.
- 'Wetsadvies W03.17.0166/II - Uitvoeringswet Algemene verordening gegevensbescherming'. Raad van State, 2017.
- Wever, Marc. 'Bezwaarbehandeling door de overheid anno 2016: Vooral vernieuwing op papier?' *Nederlands Juristenblad* 2016, nr. 44.
- . 'De bezwaarprocedure: Onderzoek naar verbanden tussen de inrichting van de procedure en de inhoudelijke kwaliteit van bezwaarbehandeling'. *Recht der Werkelijkheid* 38, nr. 2 (November 2017).

- Widdershoven, Guy. *Ethiek in de kliniek: hedendaagse benaderingen in de gezondheidsethiek*. Boom Uitgevers, 2000.
- Widdershoven, Rob. 'Een ervaring als staatsraad-generaal: op zoek naar een rechtsbeginsel'. In *De conclusie voorbij. Liber amicorum aangeboden aan Jaap Polak*, edited by M Bosma, B.J. van Ettekovén, O van Loon, H.G. Lubberdink, J.C.A. de Poorter, en B.J. Schueler. Ars Aequi, 2017.
- Widlak, A.C. 'Kan de overheid haar fouten corrigeren? #11'. *Stichting Kafkabrigade* (blog), 24 November 2018. <https://kafkabrigade.nl/home/publicaties/columns/-11-kan-de-overheid-haar-fouten-corrigeren-#idMxhU9NsiVhbrqHjaq0DtEQ>.
- Widlak, A.C., en R. Peeters. *De Digitale Kooi: (on)behoorlijk bestuur door informatiearchitectuur*. Boom Bestuurskunde, 2018.
- Wieringa, Maranke. 'What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability'. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1–18. FAT* '20. Barcelona, Spain: Association for Computing Machinery, 2020.
- Willems, D., en M. Hilhorst. *Ethische problemen in de huisartspraktijk*. Practicum Huisartsgeneeskunde. Bohn Stafleu van Loghum, 2016.
- Willems, Dick. 'Bewijzen, weten, en begrijpen. Drie vormen van kennis in de zorg'. In *Komt een filosoof bij de dokter*, edited by Maartje Schermer, Marianne Boenink, en Gerben Meynen. Boom Filosofie, 2013.
- . 'Family Medicine'. In *Encyclopedia of Global Bioethics*, edited by Henk ten Have, 1–10. Cham: Springer International Publishing, 2014.
- Willems, Dick L. 'Ethiek en de huisarts'. *Bijblijven* 32, nr. 3 (1 April 2016): 130–41.
- Willems, D.W., R. Vos, G. Palmboom, en P. Lips. 'Passend bewijs. Ethische vragen bij het gebruik van evidence in het zorgbeleid'. Signalement. Centrum voor Ethiek en Gezondheid, 2007.
- Williams, Bernard. *Truth and Truthfulness*. Princeton University Press, 2002.
- Williams, Patrick, Adi Kuntsman, Emeka Nwankwo, Danella Campbell, and Leah Cowan. 'Surfacing Systemic (In)Justices: A Community View'. *Systemic Justice*.
- Wit, Laura de, Aartjan Beekman, Christiaan Vinkers, Otto Maarsingh, en Henriëtte van der Horst. 'Antidepressiva in de dagelijkse praktijk'. *Huisarts & Wetenschap* 2019, nr. 12. Consulted 19 November 2022.
- Witteveen, W J. 'Kafka en de verbeelding van bureaucratie', *RegelMaat* 2010, afl.4
- Wolswinkel, C.J. 'Transparantie en openbaarheid: preadviezen 2022'. *VAR*, 2022.
- Wright, Shelley. *International Human Rights, Decolonisation and Globalisation: Becoming Human*. London: Routledge, 2001.
- 'Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal'. Amnesty International, Consulted 26 October 2021.
- Yeung, Karen. "'Hypernudge': Big Data as a Mode of Regulation by Design". *Information, Communication & Society* 1, nr. 19 (2016).
- Zerilli, John, Alistair Knott, James Maclaurin, and Colin Gavaghan. 'Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?' *Philosophy & Technology* 32 (2019): 661–83.

Zijlstra, S.E. 'Voorwaardelijke opzet van de wetgever: enkele kanttekeningen bij het preadvies van M. Scheltema'. Nederlandse Vereniging voor Wetgeving, 2021. <https://www.nederlandseverenigingvoorwetgeving.nl/wp-content/uploads/2021/01/Reactie-Sjoerd-Zijlstra.pdf>.

ZonMw. 'Achtergrondstudies zelfbeschikking in de zorg'. Evaluatie Regelgeving. Den Haag, 2013.



Fundamental legal explanation rights are seen to be in peril because of the use of inscrutable computational methods in decision making across important domains such as health care, welfare, and the judiciary. New technology-oriented rules are created in response to this, and human explainers are tasked with re-humanizing automated decisional processes. By providing explainees with meaningful information, explainers are expected to help protect decision subjects from AI-infused harms such as wrongful discrimination and underinformed, perilous participation in decision processes.

De Groot questions these legislative approaches in light of the longevity of many harms that are ascribed to the use of modern 'AI.' If explanation has a role to play as a tool against what can be described as knowledge related wrong-doing, law has something to answer for since its explanation rules have thus far underserved those in less privileged societal positions, before and after decisions were automated.

To conduct this critical questioning the thesis approaches explanation as a form of knowledge making. It builds a 're-idealized' model of explanation duties based on values described in the philosophical fields of epistemic justice and injustice. Starting from critical insights with regard to responsibly informed interaction in situations of social-informational inequality, the model relates duties of explanation care to different phases of an explanation cycle.

The model is applied to analyze the main explanation rules for administrative and medical decision making in The Netherlands. In 'tech-reg' discussions, both domains are appealed to as benchmarks for the dignified treatment of explainees. The analysis however teases out how the laws ignore important dimensions of decision making, and how explainers are not instructed to engage with explainees in ways that allow to fundamentally respect them as knowers and rights holders. By generating conceptual criticism and making practical, detailed points, the thesis demonstrates work that can be done to improve explanation regulation moving forward.

Aviva de Groot came to academia with backgrounds in cabinet making, filmmaking, and legal aid. She obtained her LLM from the Institute for Information Law at the University of Amsterdam. The thesis research was conducted at the Tilburg Institute for Law, Technology, and Society where she continues her research in the field of AI and Human Rights.