



UNIVERSITY  
OF  
JOHANNESBURG

## COPYRIGHT AND CITATION CONSIDERATIONS FOR THIS THESIS/ DISSERTATION



- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- NonCommercial — You may not use the material for commercial purposes.
- ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

### How to cite this thesis

Surname, Initial(s). (2012). Title of the thesis or dissertation (Doctoral Thesis / Master's Dissertation). Johannesburg: University of Johannesburg. Available from: <http://hdl.handle.net/102000/0002> (Accessed: 22 August 2017).

**What it Takes to Successfully Ground Rationality Norms  
in a Constitutive Account of Agency**

by

Danielle Swanepoel

920309863

Submitted in fulfilment of the requirements for the degree

**Doctor of Philosophy**

in the

**Department of Philosophy**

UNIVERSITY  
of the  
JOHANNESBURG

**Faculty of Humanities**

at the

**University of Johannesburg**

supervised by

**Professor Veli Mitova**

**October 2019**

*Dedicated to my mother, Elizabeth (1952-2010). This is it.*



UNIVERSITY  
OF  
JOHANNESBURG

## Acknowledgements

I cannot begin to express my sincerest gratitude to Veli Mitova who has, for all intents and purposes, pulled me through this PhD process – holding my head above water when I came close to drowning; pulling me along when I felt like all was lost; and laughing along when some of my ideas bordered on the ludicrous. Her strong leadership, unwavering guidance, immense patience, unprecedented wisdom, and great sense of humour made this PhD possible. Thank you, Veli.

I gratefully acknowledge the philosophy department (University of Johannesburg), in particular, Alex Broadbent, Thaddeus Metz, Catherine Botha, Raphael Winkler, and Hennie Lotter for providing me with helpful comments, interesting and helpful annual reviews of my work, and other valuable support.

I greatly appreciate those at SolBridge, in particular Joshua Park, and the Asian debate community, especially Loke Wing Fatt, for throwing me into the deep-end of the debating world which has undeniably contributed to my being a more strategic thinker and a better philosopher. Further, many thanks to Jamoliddin Ergash-zoda for carrying the load in the final stages of this project.

I am indebted to my friends, in particular Blaire and Sian, who provided unending emotional support for when things got rough and intellectual support when I most needed a soundboard.

Furthermore, my deepest appreciation goes out to my family. To my father, Danie, who has emotionally, intellectually, and financially supported me through many years of study – I will be forever grateful. To my brother Hein, who has always motivated me to push through the difficult times. To my brother Niel, his infectious laughter and wonderful sense of humour has brought me much joy. To my brother Fritz, my sister-in-law Erna, and my two beautiful nieces, Meah and Livia – your smiles, your openness, your heartfelt support and your love is what I hold most dear.

# Abstract

Current accounts of normativity struggle to explain the source of the normative grip of rationality norms. They fail to explain, that is, why one should be committed to adhering to rationality norms rather than to violating them, or why one should be guided by rationality norms at all. A popular fix is to appeal to a constitutive account of agency: the normative grip of rationality norms can be found in constitutive features of agency, such as beliefs, desires or drives. If accepting rationality norms is constitutive of this feature of agency, then, in virtue of being an agent, we are guided by rationality norms. This is one promising way to pin down the source of the normative grip of rationality norms. The problem, I argue in the first half of this dissertation, is that existing accounts do not actually achieve this. This failure can be attributed to one core issue: none of the accounts can explain why we should care to be agents. The purpose of my project is to provide a new account of agency which can successfully answer this core concern, and thus ground rationality norms.

In the second half of the dissertation, I propose a new constitutive account of agency which has two main tenets. First, it is plausible to think of agency as something we have *chosen* and continue to *choose*. I call this *elective agency*. It places us in a unique position to answer the question of why we should care to be agents rather than non-agents. Second, the distinguishing mark of elective agency is something I call the *worth-drive*. By embracing the worth-drive, we seek to establish the value of being agents, and of constantly re-electing agency. I argue that accepting rationality norms is constitutive of the worth-drive, and that the worth-drive is partially constitutive of agency. Being able to choose agency means that we choose to accept the rationality norms that come along with it. Because we have this constitutive drive to establish value, we care to be agents and we care to follow rationality norms. I argue that this constitutive account is the most promising for grounding the normative grip of rationality norms.

# Table of Contents

<b>Chapter 1 – Introduction</b>	<b>1</b>
1.1 The Problem	1
1.2 Background	7
1.3 Outline	12
Chapter 2: The Shtest	12
Chapter 3: Constitutive Features: Actions or Beliefs?	14
Chapter 4: Constitutive Features - Drives?	15
Chapter 5: Elective Agency	16
Chapter 6: Elective Agency and the Shtest	17
Chapter 7: The Worth-Drive	17
Chapter 8: Conclusion	18
1.4 Caveats	18
<b>Chapter 2 – The Shtest</b>	<b>21</b>
2.1 Introduction	21
2.2 Where and What is This Normative Authority?	24
2.2.1 Normative Grip?	24
2.2.2 Morality Norms versus Rationality Norms	25
2.3 Grounding Rationality Norms in Agency	29
2.4 Three Poles	31
2.4.1 Internalism and Externalism about Reasons	33
2.4.2 The Neo-Humean View	34
2.4.3 The Neo-Aristotelian View	35
2.4.4 The Neo-Kantian View	36
2.4.5 The Metaphysics of Constitution	37
2.4.6 Catch-up	38
2.5 The Shtest	39
2.5.1 Introduction to the Shtest	40
2.5.2 Non-arbitrariness	42
2.5.3 The Problem of Self-vindication	47
2.5.4 The Unavoidability of Agency	48
2.5.5 Norm-violation	51
2.5.6 The Shtest Summed Up	60
2.6 Objections	61

2.6.1 A Misconstrual of the Unavoidability of Agency .....	61
2.6.2 Should We Take Enoch’s Question Seriously? .....	62
2.7 Conclusion.....	66
<b>Chapter 3 – Constitutive Features: Actions or Beliefs? .....</b>	<b>67</b>
3.1 Introduction .....	67
3.2 Korsgaard: Self-Constitution as Constitutive of Agency .....	68
3.2.1 The Unavoidability of Agency .....	74
3.2.2 The Problem of Self-vindication .....	76
3.2.3 The Problem of Non-arbitrariness .....	76
3.2.4 The Problem of Deliberate Norm-violation.....	78
3.2.5 Korsgaard has failed the Shtest .....	87
3.3 Railton: Belief as Constitutive of Agency .....	88
3.3.1 The Shtest .....	90
3.3.2 The Problem of Norm-violation .....	90
3.3.3 Unavoidability of Agency .....	94
3.3.4 The Problem of Self-vindication .....	95
3.3.5 The Problem of Non-arbitrariness .....	97
3.4 Conclusion.....	100
<b>Chapter 4 – Constitutive Features: Drives? .....</b>	<b>102</b>
4.1 Introduction .....	102
4.2 The Shtest and its Demands .....	102
4.3 Velleman: The Drive for Sense-making.....	103
4.3.1 The Problem of Norm-violation .....	110
4.3.2 The Unavoidability of Agency .....	112
4.3.3 Non-arbitrariness .....	117
4.3.4 The Problem of Self-vindication .....	120
4.3.5 To Tally Up .....	123
4.4 The Diagnosis.....	124
4.4.1 Feature 1: Norm-violation .....	125
4.4.2 Feature 2: Unavoidability .....	125
4.4.3 Feature 3: Self-vindication .....	126
4.4.4 Feature 4: Non-arbitrariness .....	126
4.4.5 The Critical Flaw: Unavoidable Agency .....	127
4.5 Is This Just an Analytic-Philosophy Quirk?.....	131
4.6 Conclusion.....	134

<b>Chapter 5 – Elective Agency</b>	<b>135</b>
5.1 Introduction .....	135
5.2 Setting the Stage.....	136
5.3 Hierarchal Voluntarism .....	137
5.3.1 What Can We do Through an Act of Will? (P1 <sup>Chang</sup> ) .....	139
5.3.2 What Happens When We Generate Reasons? (P2 <sup>Chang</sup> ) .....	143
5.3.3 Why and How do We do This? (C <sup>Chang</sup> ) .....	144
5.3.4 The Take Away .....	146
5.4 Setting the Foundation for Elective Agency .....	152
5.5 Step 1: High-level Choice is Value-conferring .....	153
5.6 Step 2: High-level Choice is Constitutive of Elective Agency .....	156
5.6.1 The Pre-Choice Situation.....	157
5.6.2 Agency or Shmagency: Running out of Reasons .....	159
5.6.3 Choosing Agency .....	160
5.6.4 Accepting Rationality Norms .....	165
5.6.5 A Quick Recap.....	168
5.7 Conclusion.....	169
<b>Chapter 6 – Elective Agency and the Shtest</b>	<b>171</b>
6.1 Introduction .....	171
6.2 The Shtest: Pass or Fail? .....	172
6.2.1 Show how Norm-violation is Possible .....	172
6.2.2 Do not Appeal to Unavoidability to Answer the Why-care Question.....	177
6.2.3 Show how Features are Non-arbitrary.....	179
6.2.4 Don't claim Self-vindication .....	180
6.3 Objections.....	189
6.3.1 First Choice?.....	189
6.3.2 Why be an Elective Agent? .....	191
6.3.3 The Elective Agent Cannot Choose to not Accept Rules .....	194
6.3.4 Do We Really need a Constitutive Account of Agency? .....	194
6.3.5 When Choosing, it does not Automatically Mean I want the By-products.....	195
6.3.6 Is Elective Agency not also Inescapable?.....	196
6.3.7 The Bleakness Objection .....	197
6.3.8 The Phenomenological Objection .....	198
6.4 Conclusion.....	199
<b>Chapter 7 – The Worth-Drive</b>	<b>202</b>



7.1 Introduction .....	202
7.2 Just What <i>is</i> the Elective Agent?.....	203
7.2.1 What is this Drive? .....	203
7.2.2 Accepting Rationality Norms is Constitutive of the Worth-drive.....	204
7.2.3 Agents are Those Things which Embrace the Worth-drive .....	209
7.2.4 We Are Elective Agents in Virtue of the Worth-drive .....	212
7.3 Value in the Worth-Drive.....	215
7.3.1 Distinction between Instrumental and Non-instrumental Value .....	215
7.3.2 How does Value-Conferral Work? .....	218
7.3.3 High-level Choice and the Worth-drive .....	221
7.3.4 Recognitional Model of Value vs. Value-conferral.....	223
7.3.5 What I Take from This .....	225
7.4 The Final Shtest.....	226
7.4.1 Accommodating Norm-violation.....	226
7.4.2 The Unavoidability of Agency .....	228
7.4.3 Non-arbitrariness .....	228
7.4.4 Self-vindication .....	229
7.5 Objections.....	231
7.5.1 No Actual Normative Grip .....	231
7.5.2 Anything Else other than the Worth-Drive?.....	232
7.5.3 Is the Worth-drive Normative or Descriptive? .....	233
7.5.4 The Parsimony Objection .....	235
7.6 Conclusion.....	236
<b>Chapter 8 – Conclusion</b> .....	<b>238</b>
8.1 Synopsis .....	238
Chapter 1: Introduction.....	239
Chapter 2: The Shtest .....	240
Chapter 3: Constitutive Features: Actions or Beliefs? .....	241
Chapter 5: Elective Agency .....	242
Chapter 6: Elective Agency and the Shtest .....	242
Chapter 7: The Worth-Drive.....	243
8.2 The Contribution .....	244
8.3 Further Study.....	244
<b>References</b> .....	<b>246</b>

# Chapter 1 – Introduction

## 1.1 The Problem

When you drove to work this morning, there is a very good chance that, if you are a conscientious driver, you stopped at all the stop signs and red lights. If you had encountered peak-hour traffic, you would have had to sit in your car and perhaps watch other drivers go about their business. At least once, if not more, you made predictions about how other drivers will behave. For instance, if the person next to you seems anxious to get moving, there's a very good chance that given a gap, he will cut in front of you. But another prediction you probably made was that if the traffic light turns red, the anxious driver would also most likely stop with all other drivers. Why is it that you stop at the red light and predict that everyone else will too?

The simple answer is that there is a standardised set of rules or norms one must follow in order to first, get a driver's license, second, to drive well, and third, to survive on the roads. There are several varieties of rules associated with driving, such as rules about how to drive (for instance, depressing a clutch to change gears), rules about how to drive legally (obeying pedestrian crossings and stopping at red lights) and rules about how to drive well (being mindful of other drivers, not cutting in front of them). These rules are shared by all drivers in a particular country, and across many countries.

Most often, drivers adhere to these rules to avoid receiving fines, being arrested or causing harm to others. Often, drivers adhere to these rules and norms because that is just what it means to be a good driver. Therefore, if an individual desires to be a driver, it is required of them to adhere to the driving-norms associated with driving – thus establishing the normative grip of the rules and norms associated with driving.

I distinguish here, two ways in which I use the phrase ‘normative grip’. In the first sense, there are rules and norms which apply to me in virtue of being a driver, and in the second sense, there are rules and norms I accept. And so, *being* a driver and fulfilling the role of a driver means that there are certain rules and norms which automatically apply to me. In addition, there are rules and norms about driving that I accept, regardless of whether I am a driver. I may accept that a stop sign means I should stop and that a green light means I should proceed. To drive well, it would seem then that, in virtue of being a driver, one would adhere to the norms and rules associated with driving. Furthermore, one would accept the rules and norms associated with driving in order to be a good driver.

But, let’s say that a person has no desire to drive well, nor be a driver at all. Perhaps the person never enters the vehicle to begin with, and then it is not required of this individual that he adhere to norms associated with driving. Further, it would seem particularly strange if he did adhere to driving-norms whilst not driving. Consider, for instance, what we would think of a pedestrian, adhering to driving-norms, was to indicate left every time they took a left turn in their path. So, many of the rules and norms associated with driving do not apply to the pedestrian. We can say that the rules and norms associated with driving have no normative grip on him. This is not a bad thing since he does not qualify as a driver if he doesn’t find himself behind the wheel of a vehicle.

Here is an example of an individual accepting a norm, but unfortunately not a norm that pertains to him as a pedestrian. This is an example of the normative grip in the second sense: where a norm is accepted as pertaining to him. But also an example where there is no normative grip in the first sense: driving-norms do not apply to him as a pedestrian.

The same can be said for any other enterprise in which an individual partakes. Think of a cardiothoracic surgeon who often performs open-heart surgery. She knows that to be the best surgeon she could possibly be, she ought to get enough sleep, she ought to eat well, and she ought to keep mentally and physically fit. If she adheres to these norms – these rules and norms which guide her on how she ought to behave as a cardiothoracic surgeon – she is far more likely to succeed, than to fail, as a surgeon. These rules and norms have a normative grip on her insofar as she wants to be a good surgeon.

I will shortly discuss a more special enterprise, that of agency. But, before I do so, let me explain what I mean when I talk about rationality norms.<sup>1</sup> So far, I have been talking about rules for an enterprise. Now, I talk about the rules I most want to focus on: rationality rules. There are particular requirements of rationality, such that if I meet these requirements, I could be considered as rational, as acting rationally, or as deliberating rationally.<sup>2</sup> These requirements are such that if I should fail to meet them, I run the risk of being irrational, non-rational or arational.

Rationality requires of individuals that they, for instance, adopt the means to their end, thereby engaging in means-end reasoning schemes. This requires, at a minimum, that there is consistency and coherence amongst a person's beliefs, attitudes, intentions, and actions (for more on this, see §4.3); that their beliefs are normally responsive to evidence, and that their beliefs are revised in light of new evidence;<sup>3</sup> and that preferences are aligned with those beliefs.

---

<sup>1</sup> I am not overly concerned about the exact parameters or contours of rationality norms, but hope to show that there are at least *some* theoretical and practical norms acceptance of which is constitutive of agency.

<sup>2</sup> Important to note here, that some theorists do not believe that there are such rationality requirements. For more on this, see Broome (1999).

<sup>3</sup> A concern here might be that we often don't feel as though we are adhering to norms, nor are we overtly aware of adhering to norms while we get on doing what we want. I believe it's not necessary to be aware of the norms all of the time and simply put, if one is in the process of determining how to achieve an outcome, one is already adhering to rationality norms. Doing what one wants is taking means to our ends – thus an instance of adhering to rationality norms.

According to Ralph Wedgwood, in order for coherence to obtain, the following must hold:

- i) The rational agent's preferences must be transitive: whenever the agent prefers A over B, and also prefers B over C, she also prefers A over C.
- ii) The rational agent's preferences must be independent: whenever the agent prefers A over B, and the truth or falsity of a proposition p is irrelevant to this preference, the agent also prefers the gamble that gives her [A if p is true and C if p is not true] over the similar gamble that gives her [B if p is true and C if p is not true].
- iii) The rational agent's preferences must be monotonic: whenever the agent prefers A over B, and also prefers the gamble that gives her [A if p is true and B if p is not true] over the opposite gamble that gives her [B if p is true and otherwise A], then likewise, for every other pair of prospects C and D such that she prefers C over D, she must also prefer the gamble that gives her [C if p is true and otherwise D] over the opposite gamble that gives her [D if p is true and otherwise C]. (Wedgwood, 2017: S76).

What would this all look like?

For example, if the cardiothoracic surgeon wishes to perform a successful organ transplant surgery, the surgeon must hold the correct beliefs that using a sterile operating theatre, having a strong operating team, using the correct surgical tools, and being in possession of a viable organ for transplant will amount to this. The surgeon proceeds to operate in a filthy operating theatre with the help of a drunk anaesthetist, using only garden tools, and an organ that is no longer viable. After all of this, if the surgeon still somehow believes that she has performed a successfully organ transplant operation, we can pointedly say that she has acted irrationally because her means do not align with the ends she has identified for herself. Similarly, theoretical rationality requires of the surgeon that she revise her belief of it being a successful surgery in light of failing to take the means to achieve the end.

Perhaps a concern here is that it is entirely possible for the surgeon to fail at being successful at surgery without actually being irrational.<sup>4</sup> Suppose that she attended a school where they

---

<sup>4</sup> Thanks to Asbjorn Steglich-Petersen for the objection and example.

only did surgeries with garden shears and drunken anaesthetists and she learnt this was the best way to do it. Some might say then that when the surgeon uses garden shears now, it's a very rational thing to be doing because this is the belief she has formed about the best tool to use when performing surgeries, so she is taking the means to her end. This may be true, but importantly, I am not confusing rationality with success here as the objection seems to suggest. When the surgeon uses shears rather than a scalpel to perform her surgery, she is still adhering to the means-end principle which is necessary – though admittedly not sufficient – for successful action. Here, the surgeon unfortunately holds a false means-end, belief which results in an unsuccessful action, but adheres to the means-end principle, which is the rational thing to do.

Now, let's think of a more special enterprise. Let's think of what it is to be an agent. An agent can be thought of as an individual who acts or has the capacity to perform intentional actions in his or her environment. To perform actions that will bring about the best outcomes for the agent, an agent needs to follow certain rules or norms associated with action. To perform actions well, an agent should be *rational*. To be rational, an agent should accept and be guided by rationality norms. The idea is that rationality norms have a normative grip on an agent just as the driving-norms might have a normative grip on the driver. But this is not always true here because the enterprise of agency differs from other enterprises, such as driving.

If you remember, driver-norms have normative grip on an individual only insofar as he wishes to be a good driver. In the same way, norms associated with surgery have normative grip on a surgeon insofar as she wishes to be a good surgeon. With these enterprises – the driver, and surgeon – the individual wishes to be good at what they do in virtue of their wanting to be those things in the first place. But, are we in the position to say the same about agency and the rules of rationality?

At first glance, there appears to be a different kind of enterprise happening with agency. So far, I have been talking about individuals who, in virtue of wanting to do something *well*, follow rules associated with that thing. But let's go back to the driver example briefly. There are driving-norms such that if the person attempting to drive did not do those things, they wouldn't qualify as a driver. Imagine a person gets into a car and starts flailing his arms as a strange version of 'driving' – however the car never moves, nor is it even running. This person is not a driver. This is a passenger. Some enterprises require that you adhere to norms in virtue of *being* that thing, not necessarily doing it well.

The distinction here rests on the difference between constitutive rules and regulatory rules. For example, the driver must depress the clutch in order to change gears. If he doesn't do this, he most certainly will not be driving anywhere. This would be a rule which constitutes what driving is. When the driver stops at red lights, or slows down when passing a school bus, he is obeying certain rules which make sure he drives well and does not harm others. This would be an instance of regulatory rules.

Similarly, agency is an enterprise that requires you to adhere to norms in virtue of *being* an agent – thus establishing a constitutive link. What does this look like? Agents are those things which perform intentional actions. Insofar as we are those things which act in the world, we need *something* to guide us through this process. Just like the driver largely depends on the rules associated with driving in order to drive and to be a good driver, agents also require a set of norms to inform our actions – rationality norms.

Some theorists (Korsgaard, 2009; Railton, 2003; Velleman, 1996) see the link between adherence to rationality norms and agency as constitutive. This means that in order to qualify as an agent, we need to adhere to rationality norms. Such that if we didn't adhere to rationality

norms, we wouldn't qualify as agents. In the same way that the person who flails their arms about is a passenger, and not a driver.

The interesting consequence of these types of constitutive accounts is that there is no convincing reason why an agent *qua agent* should accept rationality norms as opposed to, say, flipping a coin before deciding, or acting. Just because we should do something in virtue of being a particular thing, it does not mean either that we will do it or that we should do it full stop. For example, just because a robber should rob in virtue of being a robber, it does not mean that he should rob, or that he will.

You may say that there is a reason enough in that accepting rationality norms and being guided by these norms will result in successful action, which makes her a better agent. This is where it gets trickier still, as there doesn't seem to be an immediately obvious reason why an agent should want to be a good one or a bad one. But more importantly for my argument, there doesn't seem to be an obvious reason why they would want to even be an agent in the first place. The explicit difference between the driver and the agent is that we may not desire to be agents, but agential-norms supposedly apply to all agents regardless – unlike driving-norms, which are relative to the desire to be a driver. So then, how are we to establish this normative grip once and for all? This question will be the focus of my project.

To get a better grasp of this problem, it will be useful to provide some background of the literature on the topic.

## **1.2 Background**

In the problems set out above, I defined an agent simply as an individual who acts or has the capacity to act with intention in her environment. Of course, this definition needs unpacking,



which I intend to do here. Discussions about agency are often tied into discussions in philosophy of action. Philosophy of action focuses on two main arguments: the standard *conception* of action and the standard *theory* of action (Schlosser, 2015). These two are not to be confused. The standard conception of action was primarily established through the contributions made by Anscombe (1957) and Davidson (1963). The standard conception of action centres around two core claims: i) actions should be explained in terms of intentionality, and ii) there is a close relationship between acting with intention and acting for a reason (Schlosser, 2015).

The standard *theory* of action is different from the standard *conception* of action in a fundamental way.<sup>5</sup> The standard conception of action is not committed to a causal or explanatory account of action. Instead, it hopes to say something about how intentions are related to action. However, the standard theory of action *does* commit to a causal or explanatory account of action. It does this by saying “very roughly, that something is an intentional action and done for reasons just in case it is caused by the right mental states and events in the right way” (Schlosser, 2015). The right kind of mental states are those which can rationalise an action. These mental states may include beliefs, desires and intentions. The theory claims that it is the intention which gives us reason to act. And it is these intentions, beliefs and desires which *cause* us to act. This theory is known as “the causal theory of action” (Schlosser, 2015).

An intentional action belonging to an agent is one where, if the intention were absent, the action itself would also be absent. This is essentially what distinguishes mere activity<sup>6</sup> from

---

<sup>5</sup> Important to note, the causal theory of action is often seen as the standard story of action and so, for ease of exposition, I’ll work with this version. My argument does not depend on accepting it.

<sup>6</sup> By mere activity I mean activity we do not necessarily have control over or activity that is performed spontaneously or instinctively. For instance, pulling your hand from a hot stove or your leg jerking whilst sleeping. For more on the distinction between mere activity and intentional action see Anscombe, 2000; Brand, 1984; Davidson, 1963, 2001; Ginet, 1990; Mele & Moser, 1994.

intentional actions. Intentional actions of an agent are of a purposive and more complex nature than mere activity given the nature of agency itself.

There are general features of agency which all constitutivists (and many others) can agree upon. First, that an agent is an autonomous thing capable of making choices which are under her control. Second, an agent is that thing which performs purposeful, intentional actions. Third, it is that thing which can reflect on its past and make plans about the future. Fourth, it is that thing which is aware that by making certain decisions and performing particular actions, it is going to face a set of consequences and these consequences are going to have either a positive or negative impact on her life.

As Luca Ferrero points out, “conflicting conceptions of agency would not disagree over statements like ‘agency is the capacity exerted when a subject acts intentionally as a result of her autonomous choice over alternatives she believes to be open to her’” (Ferrero, 2009, p. 320) and “the capacity to shape one’s conduct in response to one’s appreciation of reasons for action and to engage in the practice of giving and asking for these reasons (both about one’s own conduct and that of others)” (Ferrero, 2009, p. 307).

The trick is to find out in virtue of what action is distinct from activity. Answering this question amounts to locating the hallmark of agency. Theorists such as Harry Frankfurt (1988) and Michael Bratman (2007) have provided accounts of agency that hinge on the notions of self-governance and realization of intentions. Both theorists push for the idea that being self-reflective is what allows an agent to both identify with and endorse the attitudes and desires that drive intentional actions. Endorsement is a key aspect to these accounts, such that if we do not, or are unable to, endorse these attitudes and desires, then any resulting actions from these attitudes and desires are not considered full-blooded, self-governed actions.

The most uncontroversial examples are instrumental standards: how an agent should act given a certain outcome the agent hopes to achieve. For example, if I am hungry, my best bet is to source some food and then consume it to sate my hunger. However, even though these normative standards prescribe how it is an agent should behave, there is a complex distinction between norms and having so called normative reasons to do something. Not all norms automatically give me reasons to do things.

For example, a social norm might be that I should not lick my plate after I'm done with my meal. But it isn't entirely clear why it is that this norm serves as a reason for me not to lick my plate. When it comes to this social norm, what I may understand is that there is something wrong, in the eyes of society, if I lick my plate, but it isn't entirely clear at what point, or why, I accept this norm as something I should be adhering to.

The current status of rationality norms is normative in the sense that “they concern the normative characteristics of beliefs or actions, such as the characteristic of being held, or done, for good reasons” (Millar & Bermudez, 2002). Rationality norms provide a framework from which we can approach decision-making and reasonable action. The normative-grip problem here is two-fold. First, norms appear, somehow, to have an objective authority over agents – but it isn't clear how; second, it's unclear how norms are motivationally effective in governing our actions.<sup>7</sup>

Here is where the constitutivist comes in. Tiffany defines constitutivism as “the view that it is possible to derive contentful, normatively binding demands of practical reason and morality

---

<sup>7</sup> Because the focus of this project is on constitutive accounts, I will not engage in conversations such as, for instance, the exchange between Boghossian and Wright, where the aim is to understand the precise nature of epistemic rules and their relationship to justification. Boghossian argues that the justification of beliefs can be grounded in epistemic principles roughly modelled on inference rules (Boghossian, 2003); rather than to argue that to “draw on inference is to follow a rule of reasoning” (Hlobil, 2014, p. 424), Wright argues that epistemic principles may be one of the most direct ways to ensure that rules are followed responsibly (Wright, 2002; 2012).

from the constitutive features of agency” (Tiffany, 2012, p.223). For instance, Christine Korsgaard (2009) argues that we can explain what it is to be a good agent by showing what it is to *be* an agent. This is a strong strategy for grounding the source of normativity in that if a constitutivist is asked: Why do we adhere to norms? The simple answer is that we adhere to norms in virtue of being an agent. So, as an agent, being governed by norms is just something you do. And so,

If there is a constitutive norm for action, then it would not make sense to challenge that norm by asking whether one should act in accordance with it, since action just is (at least in part) behavior that falls under the authority of this norm. (Silverstein, 2015, p. 1128)

However convincing this strategy may sound, it has come under attack, most notably by David Enoch (2006). Enoch argues that even if we do accept the idea that the source of normativity can be grounded in what is constitutive of agency, it all seems to fall apart if we ask one very important question: *why should we care to be agents?*

If a [constitutivist] theory is going to work for agency, then, it is not sufficient to show that some aims or motives or capacities are constitutive of agency. Rather, it is also necessary to show that the “game” of agency is one we have reason to play, that we have reasons to be agents. (Enoch 2006, p. 186)

The problem Enoch further identifies is that if we are looking for a reason to be an agent, we are, by asking this question, appealing to norms to provide us with this reason, rendering the constitutive account essentially powerless, because we cannot appeal to something which is apparently a constitutive feature of what we already are. Therefore, “normativity will not come from what is constitutive of action” (Enoch 2006, p. 171).

I believe Enoch’s challenge is justified and leaves most constitutive accounts of agency in an uncomfortable position. I do so by considering three such prominent theories (Korsgaard, 2009; Railton, 2003; Velleman, 2004). But, I argue, we can avoid these positions and salvage the

good insights from these views if we re-evaluate some of the core assumptions upon which most of their arguments are founded. I argue that one critical assumption these accounts make about agency, which gives Enoch's criticism such force, is that agency is an *unavoidable* enterprise and that it is something essentially *involuntary*. I argue that as long as constitutivists are committed to the unavoidability of agency, they are unable to answer why it is we should care to be agents.

I have provided here a general backdrop of the project. Now, I offer a more detailed outline of what this thesis will do.

### **1.3 Outline**

In order to establish what it takes to successfully ground rationality norms in a constitutive account of agency, I argue, as a central notion, that we should abandon the assumption that agency is unavoidable. I propose a new constitutive account of agency I call *elective agency*. The foundational premise thereof is that agency is something we choose. My argument develops in two stages – a negative and a positive. The first stage (chapters 2-4) is to consider three prominent constitutivists' accounts by way of clearing ground for my substantive contribution. The second stage (chapters 5-7) is to propose a new constitutive account of agency and to show how this account is better able to answer the why-care question than any of the other accounts could.

### **Chapter 2: The Shortest**

This chapter serves to provide some important background and contextualisation of the problems associated with grounding the source of rationality norms. It begins by distinguishing different approaches to normativity and isolating the one I choose to focus on - grounding the source of rationality norms in constitutive accounts of agency. Next, I discuss the distinction

between internalist and externalist perspectives on how rationality (and other) norms have the grip they do on our decision-making processes and actions. Constitutive accounts of agency are usefully categorised in three groups: neo-Humean, neo-Aristotelian, and neo-Kantian. I argue that regardless of the groups in which an account of agency falls, they all suffer one essential failure, most notably captured by Enoch.

The failure is this: if constitutive accounts are going to push that a constitutive feature of agency enjoys a privileged normative-status, then to truly show how this constitutive feature has the normative grip required to achieve this privileged normative-status, a constitutivist needs to answer why it is that we should care to have this feature as opposed to any other feature. It is not enough to argue that because a feature is constitutive that it is non-arbitrary. We need to further show why it is that we should care to be agents rather than ‘shmagents’ or non-agents in the first place.

Under the umbrella of the *why-care question* (Mitova, 2016), one can find four problems (three from Enoch and one of my own) that a constitutivist must overcome to answer the why-care question in a satisfactory manner, and hence pass what I will call ‘the Shtest’.

The first is the *problem of norm-violation*. For an account of agency to truly show why an agent should care to adhere to a norm, the account must indicate how it is possible for her to violate this norm.

The second problem is what I will call the *problem of the unavoidability of agency*. For an account to show that agents should care, it cannot appeal to the notion that agency is just something we find ourselves in and so we should automatically care.

The third problem is what I will call the *problem of non-arbitrariness*. A theory of agency would need to show how the constitutive feature of agency is distinct from other non-agential features. Further, that this constitutive feature enjoys a privileged normative-status.

Finally, there is the problem of *self-vindication*. A constitutivist should not say that by asking (or answering) the why-care question, we are using agency to do so, and so agency is self-vindicating.

### **Chapter 3: Constitutive Features: Actions or Beliefs?**

In this chapter, I delve deeper into two popular constitutive accounts of agency. The aim is to determine how each fare with regards to the Shtest.

I first discuss Korsgaard's account and argue that it does not fare very well. First, it does not illustrate how norm-violation is possible without either completely diminishing agency or foregoing agency entirely. Second, Korsgaard wholeheartedly commits to the unavoidability of agency by claiming that agency is something we are essentially condemned to. Third, she fails to show how what is constitutive of agency is non-arbitrary and thus is unable to explain how the constitutive feature has a privileged normative-status. Fourth, Korsgaard falls into the trap of self-vindication: in the act of asking the question of why-care, we are acting, which is something constitutive of what it means to be an agent.

Peter Railton fares a bit better than Korsgaard. He shows that norm-violation is possible, but unfortunately only in a weak sense. Second, Railton also commits to the unavoidability of agency, but unlike Korsgaard, has a better explanation for it. Third, Railton shows, to a degree, how the constitutive feature of agency is non-arbitrary and thus can explain, at least somewhat, the privileged normative-status of this constitutive feature. Finally, and unfortunately, Railton also falls hard against the problem of self-vindication.

## Chapter 4: Constitutive Features - Drives?

This chapter focuses on the constitutive account of agency proposed by David Velleman. I also briefly discuss a second constitutive account forwarded by Katsafanas. What differentiates these accounts from the two explored in chapter 3 is that the present two centre around motivation. Both accounts establish a special desire or drive that they argue is the distinguishing constitutive feature of agency. Velleman's account argues that agency is partly constituted by a drive for sense-making, and that accepting rationality norms is constitutive of the drive for sense-making. Katsafanas argues that the drive for power is the constitutive feature of agency.

Both accounts fail the Shtest, I argue, but Katsafanas more so than Velleman. In both accounts, norm-violation is possible, thus passing the first condition of the Shtest, but Velleman provides a stronger case as to how this is possible. However, both Velleman and Katsafanas run up against the problem of unavoidability. Velleman argues that an agent cannot dissociate herself from her autonomy and self-understanding – two essential elements of agency. Katsafanas argues that the drive for power is inescapable. For these reasons, neither can overcome this problem.

Third, Velleman does fairly well when explaining how the proposed constitute feature is non-arbitrary by arguing that the drive for sense-making is the hallmark of agency. Finally, both Velleman and Katsafanas suffer from the problem of self-vindication, but I show that Velleman at least provides a better reason than the previous accounts as to why this is permissible. This chapter ends with a diagnosis: all accounts discussed so far fail due to one critical flaw, which is that they buy into the unavoidability of agency and treat it as a virtue of their accounts. I argue that this is where the failure lies. How so?



The flaw of all constitutive accounts discussed in this project is that they in effect derive self-vindication from unavoidability. The problem with this is that as long as you are always linking these two features, it becomes increasingly impossible to explain non-arbitrariness and norm-uniqueness.

## **Chapter 5: Elective Agency**

Given the pitfalls of the constitutive accounts of agency I have examined thus far, this substantive chapter will develop a new constitutive account of agency that does not suffer the same failures as the others. I call this account *Elective Agency*. I do this by first introducing a hybrid-voluntarist theory proposed by Chang (2009), which I use as the foundation upon which to build my account. She (2009) proposes a ‘hierarchal voluntarism’, where normativity is dependent on an act of will, when (and only when) non-voluntarist reasons have been depleted.

Hierarchal voluntarism pushes for the following claims: i) by taking something to be a reason, one is endowing it with the normativity of a reason; ii) there are instances when reasons run out and this occurs when one has sufficient reason to choose any option or when the reasons are on a par; iii) when reasons have run out, there is a way to will a consideration to be a reason – creating normativity through this act of willing.

I then argue that we are in the same situation when asked to provide reasons to be an agent rather than a shmagent, as Enoch asks us. The problem is that we have equal reason, as things stand, to choose either agency or shmagency. To argue that agency is just something we happen to find ourselves doing (as the constitutivists do) further reinforces Enoch’s concern that we don’t have reason to care to be that thing. To overcome this problem, I argue that we create reasons to choose agency over shmagency through an act of will. Because of this, I will a consideration to be a reason for me to choose agency, imbuing that reason with

normativity. My account thus overcomes the problem of unavoidability, and *a fortiori* the critical flaw in the other accounts of linking too closely self-vindication with unavoidability.

## **Chapter 6: Elective Agency and the Shtest**

In this chapter I analyse the black-box notion of elective agency in terms of the design constraints set out in the Shtest. I show that the account of elective agency is able to meet all four constraints and is thus could serve as an account to ground the source of rationality norms. I show how we can meet the norm-violation constraint by showing that when we choose agency, we confer value onto that which we choose and that which is associated with it. If accepting rationality norms is constitutive of agency, then we choose those norms. It becomes a choice for me to validate my decision to choose agency by adhering to the norm – catering for norm-violation.

Second, the unavoidability of agency is not an issue because under the account of elective agency, agency is completely avoidable. Third, by breaking the intimate link between the unavoidability of agency and self-vindication, I show that even if self-vindication happens under the account of elective agency, it's not something which should concern us. Finally, showing how these features which are constitutive of elective agency is non-arbitrary is easier to do with this account as choosing agency is the highest-level of choice, I argue. There is certainly nothing arbitrary about that.

## **Chapter 7: The Worth-Drive**

In this chapter, I provide a substantive account of elective agency by arguing that the hallmark of agency is what I call the worth-drive. The first thing I do is establish what the drive is. It motivates us to aim our decision-making and actions at identifying and establishing value. I

then argue that an essential element of elective agency is that we embrace the worth-drive and that this is essentially what sets us apart from non-agents or shmagents. I proceed to show that a constitutive aspect of embracing the worth-drive is that we accept rationality norms. The worth-drive is unable to function without accepting rationality norms and thus, I establish the necessary link between the two.

Because the worth-drive aims at establishing value, I need to offer an account of value. I explore two: one constitutive and the other not. The first sees action as value-conferring; the second treats actions as responding to value. I show that value-conferral is best suited for my purposes of showing the relationship between accepting rationality norms as constitutive of the worth-drive.

Finally, as action is value-conferring, it can also be value-diminishing. Because of this, the role of the worth-drive is to motivate us to choose actions and make decisions in accordance with what it is that validates the choice we make to be agents. Essentially, we want to be able to say: if this is what it means to be an agent, then this is something I want to be doing. From this, I proceed to show how a constitutive account of elective agency, in virtue of embracing the worth-drive, is better able to meet the problems set out in the Shtest.

## **Chapter 8: Conclusion**

In the conclusion I provide a synopsis of my argument with reference to each chapter. I also make it clear as to what the unique contribution is of this project: providing a new constitutive account which is able to better sustain the normative grip of rationality norms. Finally, I provide possible ways in which this new constitutive account can be used in future studies.

### **1.4 Caveats**

Of course, as with any research project, there are at least two core debates about rationality in which I will not be engaging here. One in particular, stands out: the idea that rationality actually has no normative grip. John Broome, for instance, argues that there are requirements of rationality but that we are wrong to think that these requirements are normative. As Broome (2005, p. 325) claims, “the view that rationality requirements are normative is the view that, necessarily, rational requirements constitute either sufficient or pro tanto reasons”.

Similarly, Niko Kolodny argues that we don't have good reason “to comply with rational requirements in general” (Kolodny, 2005, p. 509). The normative grip of rationality does not provide reasons for actions, he maintains, but instead derives from the reasons we already have.<sup>8</sup> I will not attempt here to combat sceptics about the normativity of rationality. My project is addressed, rather, to those who think that at least some rationality requirements are normative. My focus, then, is on working out how an account of agency might explain this normativity. Agnostics about the normativity of rational requirements are welcome to hear my conclusion as a conditional: *if* agency is to explain the normativity of rationality norms, then elective agency with its constitutive worth-drive is our best shot at such an account.

The second area I am not going to focus on are accounts aiming for unified theories of rationality. One example is that proposed by Robert Nozick (1993), who makes the important claims that principles very likely serve as the core of a theory of rationality and that rationality is evolutionary. Another example is Robert Audi (2001; 2004) who hopes to unify theoretical and practical rationality in a theory of experience. A third example is Jose Luiz Bermúdez (2009) who explores a possible account of a unified theory of rationality by examining – and rejecting - decision-theory as the core of such an account. As interesting as these accounts are,

---

<sup>8</sup> Denying that rationality can explain reasons for actions is explored by other theorists such as Dancy (2000), Raz (1999) and Scanlon (1998).

discussing them properly in the limited space I have here would not allow me to explore deeply enough the central topic of this project.

The third area I am not going to focus on is the question of whether we should even take Enoch's challenge seriously. There is a concern that the question of why we should care to be agents is an incoherent question to ask.<sup>9</sup> I briefly respond to this concern in §2.6.2, but addressing it properly would take a whole dissertation on its own, and a very different one from this one. This one is addressed to those who think Enoch's challenge is coherent and important.



---

<sup>9</sup> Thanks to Jessica Lerm for this concern.

## Chapter 2 – The Shtest

### 2.1 Introduction

Normativity is not a term we often use, yet norms pervade our daily lives. They inform our thinking, our actions, our decisions, telling us how we *ought to be* with respect to almost every aspect of our lives from lofty ones like being morally good to trivial ones like which fork to use when eating fish. You'd think, given the magnitude of its role, we would have figured out where normativity of various kinds comes from, or at least why we feel so strongly about its application, relevance and existence. Essentially, there are *things* which act, think, and make decisions and determine their kind of existence based on norms that don't seem to have an agreed-upon genesis – at least not in the teleological sense. Effectively, we have a donkey (actions, beliefs, decisions, deliberations etc.) and blind-folded philosophers attempting to pin the tail (normativity) – blind-folded because the source of the normative authority of the norms that govern us remains elusive.

Think of a child who runs around the garden with a pair of garden shears in her hands. Her father sees this, and immediately tells the girl to stop running with the shears. But the child is curious and asks why she shouldn't run with the shears. Perhaps her father is impatient and replies that she just shouldn't. This child is obstinate and whines “but why shouldn't I?”. Her father is now becoming more impatient and explains that it's dangerous to run with shears and that she might hurt herself; that's why she should not run with shears. The child doesn't like this answer and asks why she should care if she falls and hurts herself. This makes the father angry and he explains that if she falls and hurts herself, he would have to pay the hospital bills. The child finds this amusing because she has no concept of money nor how serious a hospital visit might be, so she says, ‘well, I don't understand money, so that doesn't affect me. And if

it's sore, it means I will get ice-cream'. Maybe at this point the father gives her a cold glare and says: 'you shouldn't run with shears because if you do, you could die'. The child is silent. This seems a good enough reason for her not to run with the garden shears.<sup>10</sup>

But even if it appears obvious that one should not run with shears, it isn't obvious why this norm should apply to *you*. It makes reasonable sense not to run with shears, but just because it is reasonable not to do this, what is it that provides this norm with its normative authority over your actions? How is it that norms can guide us and serve as reasons to perform one action over another?<sup>11</sup> That you just *should not* run with shears seems obvious enough. The reasons for not running with shears are aplenty. But why should those reasons be enough for me to *choose* not to run with shears?

Not only do those reasons inform my choice, but they further inform the way I see and deliberate about problems. The question really is, how do we get from there being good reason not to run with shears to that being the reason which moves me to act, and further, that my actions reflect consideration of this norm? Simply put, how do we get from the fact that I ought (have a good reason) not to run with shears, to my being guided by that ought (good reason)? To answer this question, we need to answer the deeper question that I have been calling the why-question: why should agents care about norms?

It seems, we have the donkey up on the board – the agent with all of her actions, thoughts, beliefs, desires and decisions – and a tail (the norms that seemingly guide the actions, thoughts,

---

<sup>10</sup> For a more detailed discussion of such iterated why's in the context of the source of practical and epistemic norms, see Railton (2003) as well as my discussion in § 3.3 below.

<sup>11</sup> I should mention that the notion of norms providing us with reasons to act has been contested by theorists who argue that reasons aren't reducible to rationality- or any other norms. Kolodny (2005, p. 511) refers to these theorists as 'nonreductivists' which include, most notably Parfit (1997) and Scanlon (1998).

beliefs, desires and decisions) that doesn't seem to fit anywhere.<sup>12</sup> Yet it just seems obvious that it *should*, because it seems to underpin everything we do! But where? One option is to ground the source of rationality norms in morality. For instance, if you were not governed by rationality norms, one could argue, you would struggle to satisfy the requirements for being a good person. The source of rationality norms can also be grounded in society. For instance, if you were not governed by rationality norms, you would potentially fail to be accepted because your actions and behaviour would be contrary to what is acceptable in society.

The source of rationality norms can also be grounded in agency. For instance, if you are not governed by rationality norms, you could potentially fail to be considered an agent. For example, the constitutive theorist argues that the authority norms have over an agent can be found in agency – thereby, promising “to reveal not merely which norms have genuine authority, but also what their authority consists in” (Silverstein, 2016).

In this chapter, I do the following: I briefly set out different approaches to normativity, concluding that constitutivism seems the most promising route to go given the focus of this project (§ 2.2-2.4). I then set up what I call the *Shtest* (§2.5).<sup>13</sup> I examine Enoch's main objections to constitutive accounts of agency, where the primary burden identified is that such accounts need to tell us why we should care to adhere to norms rather than *shnorms*, and further, why we should care to be agents rather than *shmagents*. I propose the *Shtest* as a litmus test, to test constitutive accounts of agency for whether they succeed in meeting Enoch's challenge. The closer the constitutive account gets to answering the concerns set out in the *Shtest*, the more likely this account will be to explain the normative grip of rationality norms.

---

<sup>12</sup> The source of normativity has been an awkward concept which is difficult to pin down, that's why there are so many debates out there as to where exactly it is that we can ground it.

<sup>13</sup> Thanks to Veli Mitova with whom we coined this term during one of our discussions.



## 2.2 Where and What is This Normative Authority?

### 2.2.1 Normative Grip?

I use the term ‘normative grip’ interchangeably with ‘normative authority’ to mean our caring to adhere to certain norms when performing actions or forming mental attitudes such as beliefs and intentions. Why should we care? I call this the ‘why-care question’. How do we go from there being a reason to do something to it’s becoming *my* reason to do something? Normative grip is the force a norm has to inform our actions. Some norms have a stronger normative grip over our actions than others. Something I explain in §2.2.1 – §2.2.2. Why is it that I think the answer to the why-care question lies in an account of this normative grip?

Simply put, the why-care question is concerned with why it is we care to adhere to norms. Arguably, if normativity has this force over our actions – or in some sense – has this normative grip I speak of, then of course it means that accounts of agency are a likely place for us to find the answer of why we care to adhere to norms as opposed to something else, say non-norms. For example, think of explaining why an object, when tossed into the air, falls to the earth. We explain this phenomenon by appealing to the ‘grip’ the gravitational force has on the object. Such that without this gravitational force, the object would likely not fall toward the earth. This example serves to show the kind of relationship I imagine partially exists between norms and actions – such, that without this normative grip, we would not, in any way, feel like we should fashion our actions accordingly.

Why should we care to adhere to norms? Let me use the example above again. I know that I can, without much fear, throw a sealed pocket of diamonds into the air, because I know that the gravitational force will ensure that the pocket of diamonds is returned to me. Of course, given the value of this pocket of diamonds, I care about the return of these diamonds very much

– but my knowledge that gravity is there and will have its grip on the pocket is enough for me to toss it in the air without much worry. Now, if gravity did not exist, or I was not absolutely certain it would exert its force on that exact spot, I would not be throwing the pocket of diamonds into the air. Just so, I care about my actions very much – how else do I get about in the world?

Norms ensure (at least certain kinds of norms) that my actions hit the mark most of the time. If the normative grip did not exist, or I was not absolutely certain it would exert its force at that exact time, I would probably not perform an action. And, if I performed an action on which norms didn't exert normative force, there's a very high likelihood that I'd regret performing the action. In some instances, one could argue that something that looks like an action, but which is not done under the guidance of norms is actually not an action at all – maybe just mere behaviour (this I explain in far more detail). Briefly, this is why I think an account of the normative grip of rationality norms is what we need to answer the why-care question. Next, I briefly look at other places where we think of normative grip.

### 2.2.2 *Morality Norms versus Rationality Norms*

When we ask questions about how one ought to behave, or how one should *decide* in any given situation, we are asking normative questions. Even if we accept that there are certain norms and principles we ought to adhere to, we require further deliberation to ascertain the correctness of these norms and principles, or at least have the capacity to decide which of these norms and principles we have most reason to accept. Often, we cite reasons for actions according to what

action is most rational to perform. Thus, rationality norms play a pivotal explanatory role, not just guiding action but also providing reason for action.<sup>14</sup>

Theorists such as LeBar (2008), Mackie (1977) and Moore (1903) argue that there is an intimate connection between morality and rationality norms, such that it is possible to establish the normative grip of rationality norms in morality. But, if this were so, there shouldn't be instances where morality norms lose their grip over an agent while, at the same time, rationality norms retain theirs.

There are instances where it is reasonable to adhere to rationality norms rather than morality norms, and so, grounding rationality norms in morality just doesn't seem like the best option.<sup>15</sup>

If there are instances where morality fails to guide us, but rationality norms still stand strong, then we shouldn't be grounding rationality norms there. Morality just does not have the explanatory power necessary to explain the instances where rationality norms still have a normative grip when morality norms don't.<sup>16</sup>

Let me use the trolley problem to explain this relationship between morality norms and rationality norms.

This thought experiment (Thomson, 1985) would have us position ourselves as a bystander on a bridge watching a runaway trolley hurtling towards five people tied to a track. Accompanying

---

<sup>14</sup> Of course, as mentioned in the introductory chapter, theorists such as Broome (1999; 2005; 2008) and Kolodny (2005) argue that reasons for actions cannot be gotten from rationality norms.

<sup>15</sup> Kant, for instance, characterises the categorical imperative as a rationally necessary principle which we ought to follow even in the face of desires or drives which contradict this principle. The role of the categorical imperative in decision-making and action is explored in *The Groundwork of the Metaphysics of Moral* (Kant, 1785); *The Critique of Practical Reason* (Kant, 1788), *The Metaphysics of Morals* (Kant, 1797), *Anthropology from a Pragmatic Point of View* (Kant, 1798), *Religion within the Boundaries of Mere Reason* (Kant, 1793). For other primary sources exploring the idea that rationality norms and morality are closely intertwined see notably Hobbes (1651), Locke (1689) and Aquinas. For more contemporary discussions, see Coleman (1990); Dancy (1993); Elster (1991); McDowell (1978); Scanlon (1998); Shafer-Landau (2003); Wedgwood (2007).

<sup>16</sup> Some theorists argue that morality norms can only be grounded in societies, as is argued by ethical relativists, most notably Harman, 1975; Prinz, 2007; Westermarck, 1960; Wong, 1984, 2006.

us on the bridge is a large enough person, that were we to push him off the bridge, he would successfully land on the tracks, stopping the train from killing the five people, but unfortunately, he would die in the process. The question that the problem is supposed to tease out is whether it is permissible to sacrifice the life of one person so that we can save five others.<sup>17</sup>

Often, this thought experiment becomes far more complex when we introduce back stories, such as the person on the bridge has the cure for cancer and the five on the railroad are serial killers. Either way, we ask an ‘*ought*’ question: what *ought* one to do in this situation?

The ‘should’ in this question is asking us to imagine a world where this is possible and to then prescribe behaviour to the person who finds himself in this unfortunate situation. Importantly, the answer to this question will not just apply to this person in this one instance. Suppose that we have determined that we should save the five lives at the price of one. Given this aim and the set-up of the experiment, rationality requires that you push the man over the bridge. Compare: you conclude you should save the five lives, then shake his hand, and take him for a coffee to the nearest café.

Sometimes norms of either kind can fail to have a grip on us. Suppose that after you concluded that you should push the large man off the bridge, I informed you that I’d forgotten to mention a tiny detail in the set-up of the thought experiment: that the large person standing next to you on the bridge is your spouse. You then have two options: you could revise your verdict that

---

<sup>17</sup> There are many other versions of the problem aiming to answer versions of this question. For instance, the classical version will have us in the position of the trolley manager of sorts – we are in the unique position to pull a lever and change the direction of the trolley so that it can hurtle down either one of two tracks. Unfortunately, on the one track along which the trolley is currently making its way, lie five incapacitated people. On the other track lies just one person. If you pull the lever, you change the direction of the trolley and it will kill the one person rather than the five (Foot, 1967).

you ought to save the five lives or you can maintain it but notice that the original verdict has lost its motivational grip on you.<sup>18</sup>

Suppose we go for the second option. Then you judge that you should push your spouse onto the tracks. But just because this is something that you judge you ought to do, it isn't clear that this is something you will *actually* be moved to do. In this instance, it still has a normative grip over me insofar as I am aware that I should save the five lives, but yet I still choose not to push my spouse over the bridge, I acknowledge that this is not the right thing to do. I have also been referring to this phenomenon as a norm failing to guide one.<sup>19</sup>

But it is possible that while morality norms fail to guide you, rationality norms succeed. You may, if this unfortunate situation should ever arise, calculate that your spouse is indeed large enough to stop the train from killing the five people. But, you are acutely aware of what your life will be like without your spouse in the picture, and worse still, that you would be responsible for his or her death. You weigh up the value of your life – as the person who pushes your spouse off the bridge and realise that your life has far less value after this act than if you did nothing instead. This proposed action you ought to take (at least according to morality norms) does not drive you to action. Instead, accepting rationality norms governs your inaction (you do not push your spouse off the bridge) providing you with practical (in this case prudential) reasons as to why you should not push your spouse off the bridge.

---

<sup>18</sup> If you revise your verdict, you are probably some form of 'partialist', believing that morality's oughts are not impartial in the way envisaged by many traditional theorists, but in part depend on one's personal relationships (e.g., Williams, 1981).

<sup>19</sup> John Mackie (1977) outlines three reasons for this. First, that morality norms vary from different cultures, second, it isn't clear what kinds of facts could make moral-norms true in the sense that we feel compelled to adhere to them? Third, even if there are moral truths, how can they be objectively known? My example is closest to the second.

Thus, when you fail to push your spouse off the bridge, it is possible that you are being governed by rationality norms, such that these norms justify your inaction – making this inaction acceptable– at least to you. This suggests that moral and rationality norms come apart, making grounding of the latter in the former not as automatically appealing as it appeared initially.

### **2.3 Grounding Rationality Norms in Agency**

The problem with grounding the source of rationality norms in morality for instance, is that the move in fact fails to explain the normative grip of rationality norms. As different as such accounts are from each other, they will always invite the further question of why we should care to be moral agents or social agents. Intuitively, it just seems obvious that we care more about adhering to rationality norms in virtue of, for example, self-interest. But more is needed than an intuitive guess.

So, what is required is that we provide a theory that grounds the source of rationality norms in a way that it is difficult to see how we could do anything else but be rational. A way to do this is to focus our attention on what it means to be an agent. An agent is that thing which performs intentional actions in an environment. To do this, it seems pertinent that we adhere to rationality norms.

Being an agent in part means identifying options which will bring about a desired outcome. This involves, at a minimum, at least two elements: that we hold correct beliefs, and that we perform correct actions (Wedgwood, 2002). These two things are at least necessary for *successful* action. If, as an agent, I intend to cross the road and I climb a tree instead, this action will not amount to my crossing the road. My goal is left wholly unsatisfied. To cross the road, it would be required of me that I intentionally place one foot in front of the other until I find

myself on the other side of the road. If I adhered to rationality norms, I would know this to be the case. There are two reasons why I might decide to irrationally climb a tree instead: 1) I am incapable of following rationality norms at that moment. Perhaps I am drunk or suffering a psychological breakdown. 2) I don't *care* to follow rationality norms. Perhaps I know that there are these norms, but I don't see how they are relevant to me, nor do I care if they are.

Agents are those things which expect that when they act, they will bring about a desired outcome. This expectation cannot be met if we just don't bother accepting rationality norms. It is debatable that the thing which intends to bring about an outcome but doesn't accept rationality norms will ever succeed as an agent. This is where the constitutivist enters the discussion. The constitutivist, recall (§ 1.2), maintains that we adhere to norms in virtue of being agents.

What makes constitutive accounts so attractive is that they promise to tell a story of why rationality norms have normative authority over our beliefs and actions. The strategy behind constitutive accounts is straightforward. As agents, rationality norms have normative authority over our actions insofar as we perform intentional actions or are believers. The idea is that because our actions and beliefs are guided by these norms, when we act or believe it is because we have accepted these norms and are guided by them. Think, for a moment, of playing a game of poker. If we obey the rules set out in the game of poker, we can be said to be playing the game of poker. Hence, if we are playing the game, it means we have accepted the rules and are guided by the rules. The moment we throw our cards down on the table and cry out "Blackjack!" is the moment we are no longer playing poker, but something else entirely.

Similarly, many constitutivists (Ferrero, 2009; Korsgaard, 2009; Railton, 2003) would argue, when we do things that agents do, it is because we adhere to rationality norms and further, are

expecting those norms to guide our deliberations, beliefs and actions. The moment we do something else that is contrary to these rationality norms is the moment that we aren't really playing the game of agency. Constitutive accounts are appealing because the argument is such that by being agents, we have already grounded the normativity of rationality norms.

Agent-constitutive accounts<sup>20</sup> of normativity try to argue that normative standards, aims or motives are constitutive features of agency. Further, that these features are the source of the normative grip or of the authority norms (such as rationality norms) have over our beliefs, desires, deliberations, motivations and actions. The outcome is that if we were not to possess these specific features of agency, we would not be agents. Consequently, if the norms or standards or aims didn't have the requisite grip on us, our agency is likewise undermined.

I think this is a promising strategy and one I choose to explore in this project. What I hope to establish is first, that despite their many attractions, current constitutive accounts fail to capture the normative grip of rationality norms (Chapters 3 and 4); and, secondly, that my constitutive account of agency will fare better than these accounts (Chapters 5, 6, and 7). But first, I'd like to take a moment to provide a brief background with regards to three different approaches to the source of normativity to situate the reader in this debate. To be clear: my aim here is not to set up arguments against them, but rather to provide a prima facie case for the comparative attractiveness of constitutive accounts of agency.

## **2.4 Three Poles**

Supposedly, every intentional action is performed for a reason. It is unclear, however, what the source of such reasons and their normativity is. For instance, in Korea, it is a norm to remove

---

<sup>20</sup> Acknowledgement to Veli Mitova who introduced me to this term in discussions we had.



your shoes before entering a home or some establishments. But, if you failed to remove your shoes, there wouldn't be catastrophic consequences such as death, maybe just a disgruntled homeowner. Knowing that the consequences are not dire, why is it that we still remove our shoes? What about that norm gives it authority over our actions?

In the case of cultural norms of this kind, you should care – and only care - insofar as you care about the homeowner's feelings or carpets, and perhaps about matters of appearance more generally. But rationality norms are arguably not like that. It seems both more and less obvious why we should care. More obvious, because they are more deeply entangled with our lives. Yet less obvious, because it proves quite elusive to pin down the actual reason for caring.

Constitutive accounts situate the normative grip of rationality at a constitutive level. The idea is that in virtue of having certain *desires*, or *believing* in a certain way, or *acting* as we do, we are able to ground the source of normativity in constitutive features of these things. For instance, normativity can be derived from our desires or what our belief is aimed at. Most constitutive accounts are grounded in one of three poles: neo-Humean, neo-Aristotelian, and neo-Kantian views.<sup>21</sup>

Here, I explain the difference between an internalist and externalist conception of reasons for actions, looking at a potential neo-Humean account, a neo-Aristotelian account, and finally a neo-Kantian theory of the source of normativity. The aim is to show that constitutive accounts take various forms and considerations. Throughout this project, I explore several different constitutive accounts of agency and they are likely to fall under one of these banners.

---

<sup>21</sup> To be clear, I use 'neo' in order to emphasise that I am not attributing these views to the authors themselves, but to the spirit of their – and followers' – writings and ideas. For good exemplars of the Neo-Humean view, see Bagnoli, 2002; Lenman, 2010, 2012; and Street, 2008, 2010, 2012. For exponents of the Neo-Aristotelian view, see Hoffe, 1993; LeBar, 2013; and Millgram, 2005. For more on the neo-Kantian view, see Korsgaard, 1996, 2000, 2009; O & Neill, 1989; Rawls, 1980, 1989, 2000, and Velleman, 2000.

#### 2.4.1 Internalism and Externalism about Reasons

When an agent performs an intentional action, in the full-blooded sense, she does so based on reasons, that is on what she thought favours her action. Further, these reasons are often good reasons for performing the action.<sup>22</sup> The former reasons are known as motivating; the latter as normative. For instance, the fact that running with shears is really dangerous is a normative reason not to run with shears. The fact that our country drives on the left-hand side of the road (depending on the country of course), gives a driver a normative reason to drive on the left-hand side of the road. One central debate in metaethics concerns how we should understand normative reasons. In particular, if we can agree that agents seldom perform intentional actions without reason, we need to show how the reasons out there – normative reasons – become the reasons for which they act – their motivating reasons.

According to internalism, in order for a consideration to serve as a normative reason for an agent to act in a certain way, the consideration must be appropriately related to the agent's motivations. One consequence of this is that the agent must be in a position to accept the consideration as a reason for action. But, as Robert Audi puts it, "reasons internalism is highly controversial" (Audi, 2003, p. 132). Audi maintains that there is a normative reason for action even if an agent does not accept the consideration. Furthermore, there can be a normative reason for an action "whether or not the action would satisfy any of one's non-cognitive motivational state" (Audi, 2003, p. 132). This is so-called externalism about reasons.<sup>23</sup>

An externalist would claim that an agent can have a normative reason for action just because there exists an externally grounded judgement, the truth of which is independent of anyone's

---

<sup>22</sup> This lumping together of reasons and norms is a mere convenience. I am well aware that many views go under the banner 'externalism', and there are many subtle and not so subtle distinctions amongst them. For a nice overview, see Finlay & Schroeder (2017).

<sup>23</sup> For more on the internalism/ externalism divide, please see Williams, 1979.

beliefs or motivations. Simply, “this kind of external grounding implies that one’s normative reasons for action – the kind that would justify it – are not grounded in anyone’s non-cognitive motivational states” (Audi, 2003, p. 132). For instance, there are actions which are considered abhorrent or universally wrong, such as premeditatedly taking an innocent life. Because this universal wrong exists externally, this fact alone provides reason for not performing these actions.

Note, that often, constitutive accounts will be either externalist in nature or internalist. The constitutive account I propose in Chapter 5, however, does not depend on this distinction and both the externalist and the internalist can accept it. This neutrality is an advantage for my account. I discuss it in Chapters 5, 6 and 7.

#### 2.4.2 *The Neo-Humean View*

Discussions regarding the source of normativity revolve around three prominent poles: the neo-Humean, the neo-Aristotlean, and the neo-Kantian approach (Cullity & Gaut, 2003, p. 3). I’ll briefly explain each of these in turn, in order to better situate the views that I discuss in the next two chapters, as well as the one that I end up defending. The neo-Humean view holds that all normative reasons are internal. One way of understanding this is by thinking of them as hypothetical in structure: the normative reasons for action an agent may have are always dependent on an agent’s motivational drive or tendencies, or - in typical Humean fashion - desires.

This view, importantly, considers desires as explaining all motivation, and conceives of these desires as essentially arational. To establish what normative reasons an agent has, we would need to examine “the nature of an agent’s arational motivational tendencies to determine the character of his reasons” (Cullity & Gaut, 2003, p. 4). A standard problem with understanding

normative reasons in this way is that, if they are dependent on arational desires, then it isn't clear what is normative about them and how they could make it the case that we ought to do something.

In order for a person, under the neo-Humean account to act rationally, a person would need to have a goal that would be best served by rational behaviour. In other words, the neo-Humean theory of rationality holds that “rationality is basically a matter of efficiency in serving one’s intrinsic ends, or goals, where a person’s intrinsic goals are taken as a given” (Railton, 2006, p. 265). As long as an agent has a desire or goal which will be well served by adhering to rationality norms, then this agent has a practical reason to act in adherence of rationality norms. So, for example, a neo-Humean agent may have a strong desire not to be arrested, which is what prevents him from stealing from the local grocery store.

This neo-Humean agent is moral insofar as being moral serves to satisfy a desire of his. Here, morality norms are purely instrumental in nature. The neo-Humean cares about following these norms as long as they promote a goal of his. Otherwise, those norms mean nothing to him. The problem with this account is that *any* norms really, as long as they serve the goal of the agent, can serve as a normative reason for action. This means, under this account, that while the motivational grip of rationality norms is very strong, their normativity is weak.

#### 2.4.3 *The Neo-Aristotelian View*

The neo-Aristotelian view, by contrast, sees normative reasons for action as dependent, not on desires, but rather on “the nature of free rational agency and specifically human nature... independently of our contingent motivational natures” (Cullity & Gaut, 2003, p. 4). Thus, this account does not see all normative reasons for action as hypothetical, but rather has room for non-hypothetical or categorical reasons. Particular to the neo-Aristotelian account is the

*recognitional view*. The recognitional view holds that “practical thought is principally thought *about* good reasons” (Setiya, 2007, p. 86). According to this view, there is a normative reality which is external to ourselves. This is in stark contrast to the neo-Humean account, on which reasons are entirely internal to us. Neo-Aristotelian rationality norms on this view are grounded in the good life.

As many have pointed out, however, “the recognitional approach is subject to a decisive problem: it relies on the claim that practical thought takes place under the guise of the good” (Setiya, 2007, p. 88). In a nutshell, the problem is this: acting under the guise of the good means that I always take myself to be acting for good reasons. This rules out both perverse (e.g., Velleman, 2000: Ch. 5) and akratic action. Such actions involve precisely acting in order to attain something avowedly not good or acting against one’s judgement that another action would be good, respectively. If it is possible that an agent can perform an action regardless of whether the agent considers those to be good reasons, then it means that thinking about the goodness of an action “cannot be identified with the standard of practical reason” (Setiya, 2007, p. 88). So, because it is possible to perform actions without consideration of the good, it is difficult to see how it is that all our actions and deliberations are aimed at the good. Because of this, we cannot establish the normative grip of rationality norms on such an account.

#### *2.4.4 The Neo-Kantian View*

The neo-Kantian view is different from the other two in many ways, but the most important for my purposes is that it considers normativity as being grounded in the constitutive features of agency. According to this view, normativity does not spring from our desires or goals, but rather is rooted in what it is that fundamentally makes us agents, contra the neo-Aristotelian approach which holds that reasons for actions can be found external to our agency (albeit

internal to what makes us flourish qua agents). The neo-Kantian view is powerful in many ways, but primarily in that rationality norms are “universally and necessarily binding for all rational beings” (Bagnoli, 2017, p. 14) in virtue of their being agents. This way of thinking about the source of rationality norms has interesting consequences. The most striking is that if an individual fails to adhere to these norms, then it is debatable if that individual still qualifies as an agent.

This view offers a far stronger theory of the normative grip of rationality norms in the following way: if you are an agent, then you are that thing which accepts rationality norms.<sup>24</sup> If you do not accept rationality norms, then you are either not an agent or your agency is somehow diminished. This view is different from the neo-Humean approach as adhering to rationality norms does not depend on our desires but rather on what is constitutive of our agency. It is similar to the neo-Aristotelean view in that it too is constitutive: accepting rationality norms is constitutive of being virtuous or of the good life. But the neo-Kantian view is different from the neo-Aristotelean in that the former is not vulnerable to the guise of the good objection, as it is not the case that all actions are aimed at the good. Instead, our actions are already rational because, as agents, we just are those things which adhere to rationality norms.<sup>25</sup>

#### *2.4.5 The Metaphysics of Constitution*

What does it mean to say that a particular feature can be constitutive of something else? In this dissertation I will not commit to any particular metaphysics of constitution. Here however is a

---

<sup>24</sup> Please note that for the purpose of this project, these are just prima facie considerations, not decisive arguments. As discussed in the introduction, those who are unconvinced should consider my overarching conclusion as a conditional.

<sup>25</sup> This is not the last time I will explore these different views. In the next couple of chapters I explore, specifically Neo-Kantian views by looking at authors such as Christine Korsgaard (2009; 1997), Peter Railton (2003), David Velleman (2004; 2000).

rough idea of the notion that I have in mind. Ylikoski (2013) identifies four features of constitution:

- a) There is a relationship of asymmetry between the constitutive part and the system it constitutes.
- b) The relationship of constitution is irreflexive in that nothing is able to constitute itself, it must constitute something else. This makes the relationship of constitution different from the relationship of identity.
- c) Constitution, and the relationship thereof, is synchronous. This means that it occurs at the same time – there is no time gap between a constitutive feature and the thing it constitutes – if there were some kind of time gap, this would be a causal relation and not a constitutive one.
- d) The relations of constitution do not exist independently.

Korsgaard (1999) provides a helpful example – that of house-building – to illustrate the notion of constitution. The constitutive standard of building a house is determined by whether the house is able to serve the function of a house – to provide shelter. If the house is built without a roof, then it does not provide shelter and so fails to count as a house.

#### 2.4.6 *Catch-up*

So far, I have discussed how we approach normativity, especially rationality norms. I explained that as agents we perform intentional actions. Intentional actions which are based on the desired completion of a goal would probably be unsuccessful if the action was not governed by rationality norms. It just seems unlikely that a person would succeed as an agent if they either *could not* accept rationality norms or if they did *not care* to adhere to rationality norms.

## 2.5 The Shtest

Constitutive accounts of agency seem, at first sight, a promising way to ground the normative grip of rationality norms. The idea is that normativity can somehow be grounded in what is constitutive of agency or action. It is then in virtue of being an agent and acting, that we already adhere to rationality norms. Simply, we adhere to norms in virtue of being agents.

The thinking can be broken down as follows:

P1. Adhering to norms is constitutive of X (X being any feature a constitutivist may argue for- be it action, desires, capacities or drives).

P2. X is constitutive of Agency.

C. Adhering to norms is constitutive of Agency.

What is especially compelling about this is that it appears, at first glance, that as long as we can show that someone is an agent, we can locate the source of their norm-adherence. The way to go about this is to show that without this constitutive feature, agency wouldn't occur; and then to show that without adhering to norms, the constitutive feature would not occur. It is, overall, an appealing approach, but suffers, to my mind, a crippling blow when David Enoch (2006) asks one very important question: why should we *care* about being agents? This project centres around successfully answering this question. I now show the importance of answering it.

Using Enoch's paper as my springboard, I introduce the Shtest. The Shtest sets up four criteria that need to be met in order to satisfactorily answer Enoch's question, and hence for a successful constitutivist account of the source of rationality norms. (Three of these criteria are



Enoch's and one is mine.) I use the Shtest as a litmus test. Essentially, I will be measuring each constitutive account I examine in chapter 3 and 4 against the Shtest. The account which is most successful is the one which can best ground the source of normativity in the constitutive account. I unpack each criterion and show how it is necessary for a successful answer to the why-care question.

### 2.5.1 Introduction to the Shtest

But first, let me briefly explain Enoch's position. He argues that "normativity will not come from what is constitutive of action" (Enoch, 2006, p. 33). The idea is this: if constitutivists hold that agents are those things which adhere to norms, then adhering to norms is something an agent should *care* to do in virtue of being an agent. This would provide us with non-hypothetical reasons for caring about norms. But, for Enoch, this is not as clear cut as the constitutivist presents it. For one, why should the agent care to adhere to norms rather than something else, like shnorms (Enoch, 2006) for instance?

We can think of shnorms as being a kind of artificial norm. It's something that looks like a norm, requiring something of us, but it clearly is not a norm. Ideally, a distinction between a norm and a shnorm would be that norms have a certain kind of oomph over deliberations and actions in the way that shnorm might not. And then, as Enoch argues, even if the constitutivist argues that adhering to norms is something an agent does in virtue of being an agent and adhering to shnorms just is something an *agent* will not do, in virtue of being an agent – this point can be further pushed. Enoch will then ask the constitutivist why one should care to be an agent rather than, let's say, a shmagent.

A shmagent is, what I imagine, an artificial agent. Something that acts like an agent, looks like an agent, but is not one. In other words, something that performs shmactions without a

constitutive aim. So, then there are two questions Enoch asks: why care to adhere to norms and why care to be an agent.

Let me say a bit more about the why-care question. Any constitutive account of agency which hopes to ground the source of normativity, and rationality norms in particular, needs to show that following norms is something that is important for the agent to be doing, and also, that this is something an agent should be caring to do. So, when constitutive accounts claim that to be an agent is to adhere to rationality norms, this does not mean that, as an agent, I should automatically care to follow these norms. It isn't clear how a feature, being constitutive, is imbued with a privileged normative-status in virtue of it being constitutive. This line of thought can be taken further with alarming consequences as has been explored by Enoch:

But why should I be an agent? Perhaps I can't act without aiming at self-constitution [or any other feature of agency], but why should I act? If your reasoning works, this just shows that I don't care about agency and action. I am perfectly happy being a "shmagent"— a nonagent who is very similar to agents but who lacks the aim (constitutive of agency but not of "shmagency") of self-constitution. (Enoch, 2006, p. 41)

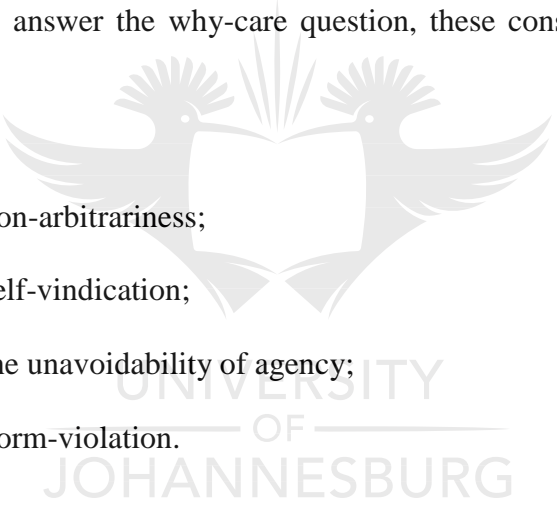
If constitutive accounts are going to push for such a strong connection between agency and rationality norms, then they fix the problem of why we should care to adhere to rationality norms, but at the cost of raising the new one of why we should care to be agents. Let me explain.

Being agents, our actions, thoughts, desires, beliefs and all else included under the realm of agency can be understood as guided by norms. Our question here is why we should accept to follow these norms in the first place. What about these norms makes them more attractive than shnorms? Enoch argues that there must be something which makes adhering to norms more attractive than adhering to shnorms. If this answer or reason cannot be provided, then the normativity we have is as valuable as shnormativity – which, by definition, is arbitrary and hence not very valuable at all. In other words, unless we can show what makes normativity

more desirable than shnormativity, it would be near impossible to show how normativity has its normative grip.

The go-to-answer the constitutivist offers is that as long as we are agents, then adhering to norms is something we do in virtue of this agency. There's no question about norms being more attractive than shnorms – we just *do* follow norms. But, Enoch argues, this isn't enough. The issue then is not about caring to adhere to norms, but rather about caring to even be agents. Unless there is enough reason to care to be agents, there is not enough reason to care to adhere to norms.

I will now argue that to answer the why-care question, these constitutive accounts should overcome four problems:

- 
- i. The problem of non-arbitrariness;
  - ii. The problem of self-vindication;
  - iii. The problem of the unavoidability of agency;
  - iv. The problem of norm-violation.

I go through each of these in turn in the sub-sections to follow.

### *2.5.2 Non-arbitrariness*

As mentioned previously, a schema for the arguments made by constitutive accounts of agency could look something like this:

P1. Adhering to norms is constitutive of X (X being any feature a constitutivist may argue for-be it action, desires, capacities or drives).

P2. X is constitutive of agency.

C. Adhering to norms is constitutive of agency.

This argument seems to cement normativity and in turn, norm-adherence in that which is constitutive of agency. If you are an agent, you *will* adhere to norms because that's just what it means to be an agent. But, if we apply Enoch's why-care question, then the constitutivist is in big trouble. Even if we accept both premises and admit that this is a seemingly fantastic argument to ground normativity, the weakness lies in getting us from the conclusion to answering the question of why we should care about such norms.

In order to reach the conclusion that we should care to adhere to rationality norms, we need to show how it is we care to be an agent. If no convincing argument can be had as to why we should care to be agents, then this constitutive argument doesn't really get us to our final destination.

Let's for this section, assume that agency is an important thing to have. A further problem faces the constitutivist: she still needs to explain why the feature she argues constitutes agency is non-arbitrary. Just because something is a constitutive feature of agency, Enoch points out, it doesn't follow that this feature is non-arbitrary (Enoch, 2006, p. 43). Yet, it is clear that until it is shown that the feature in question is non-arbitrary, it is in no position to ground the normativity of anything.

First, just because something is a constitutive feature, does not suddenly explain why I would want to adhere to norms rather than shnorms. Maybe I just don't care that just because a certain feature is constitutive of my agency, and just because this feature is somehow imbued with

normativity – that it should matter to me at all. Second, even if we buy into the idea that constitutive features of agency are somehow non-arbitrary,

Intuitively speaking, pleasures and pains – [...] seem much less arbitrary and much stronger candidates for normative significance than any motives or capacities, plausibly considered constitutive of agency. (Enoch, 2006, p. 44)

We'd need a really good explanation as to how these features are distinctly different from pleasures and pains – such that they have a normative grip over our actions in a way that pleasures and pains do not.

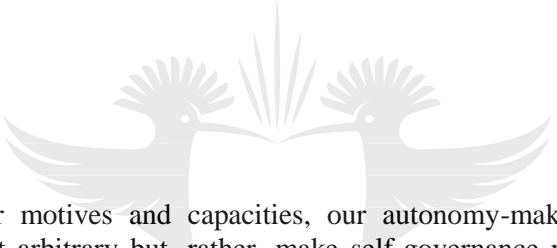
If the idea is that adhering to norms is constitutive of a particular feature, then we would need to show that this feature is what is constitutive of agency – only then could we perhaps say that that feature in particular has privileged normative-status. But if it is discovered that the same feature constitutive of agency is actually also a feature of say a mechanical clock, and clocks clearly don't adhere to rationality norms, then that feature cannot be imbued with a normative privileged status. Or, to give another example, if the constitutive feature of agency is a common desire or attitude all creatures have – such as the desire to have food or the desire to sleep. We would need to show how this constitutive feature of agency is non-arbitrary, it cannot be an ordinary feature shared by all things (given that it's supposed to be distinctive of agency) and it cannot be a feature like all our other features.

Korsgaard, for example, argues that actions are constitutive of agency, and because actions are always aimed at the good, this provides an explanation of the source of rationality's normativity. Velleman argues that there is a particular drive which is constitutive of agency and because this drive is always aimed at sense-making, this provides a sufficient account of the source of normativity. Other theorists make similar moves. As I will discuss shortly, Peter

Railton argues that belief is constitutive of agency, and that because beliefs are propositional attitudes aimed at truth, this provides a sufficient account for the source of normativity.<sup>26</sup>

The issue in all cases, as discussed at length by Enoch, is that it is unclear how these propositional attitudes, actions, drives or other such mechanisms provide a sufficient account of the source of normativity. We are not in the position to identify what it is about these propositional attitudes, actions, drives or mechanisms which makes them unique from other propositional attitudes, actions, drives or mechanisms which do not have this normative grip. A constitutivist response may be that what makes these features unique is that they are *distinctive* features of agency.

According to Rosati<sup>27</sup>



Unlike our other motives and capacities, our autonomy-making motives and capacities are not arbitrary but, rather, make self-governance possible: they are motives and capacities without the effective operation of which we would not be agents and evaluators at all. (Rosati, 2003, p. 522)

But, as Enoch points out, it doesn't help to say that these attitudes, actions, drives or mechanisms are unique to agency and thereby hold some kind of normativity-conferring status. Rosati argues that there are essential features of agency, in particular, "the capacity for critical reflection" and "certain cognitive capacities and motives that specially equip [us] to function as agents" (Rosati, 2003, pp. 511-512). The capacity for reflecting critically is the ability an individual has to 'step back' and evaluate their desires, motives etc and to decide if this is the sort of person they wish to be. Essentially, the critically reflective individual has in mind an ideal self, which is to say, a person which is in possession of complete rational capacities and

---

<sup>26</sup> But see, e.g., Steglich-Petersen (2006) for a convincing way of blocking such moves from the aim of belief to normative conclusions of this kind.

<sup>27</sup> I keep my discussions on Rosati short and do not focus on her in this project. Primarily because her account is vague regarding the constitutive feature of agency.

full information about the scenario in which they find themselves. This notion of the ideal self is what guides their desires and motives, such that, if done well, will align the actual self with the ideal self.

For an individual to be able to critically reflect, they require, at a minimum, the right kind of motives and capacities to do so. Rosati refers to these as “autonomy-making capacities and motives” (Rosati, 2003, p. 512). She does not give us a fully-fledged picture of these capacities and motives, as she is only concerned with spelling out how reflecting on them serves to partly constitute our agency (Rosati, 2003, p. 513). Instead, she says:

consider, first, that our autonomy-making motives and capacities make normative inquiry and action possible. Without them, we wouldn't be capable of raising and answering our normative questions; indeed, we wouldn't have normative questions or, for that matter, normative views or values. (Rosati, 2016, p. 194).

The problem should be clear. Anyone making these kinds of claims about the uniqueness of constitutive features would need to show how something being unique to agency also amounts to its being able to confer normativity. So just saying that a feature is distinctive of agency, does not amount to showing how it can ground the normative grip of rationality norms.

Arguably, we should not run non-arbitrariness and uniqueness together.<sup>28</sup> It is possible to think of entities which share certain features but where these shared features can still be non-arbitrary. For example, sex is a feature which is not unique to agents, yet has a clearly non-arbitrary status. And so, this calls into question the need for a uniqueness criterion. But what does this have to do with the why care question? Enoch neatly draws out the connection between uniqueness and non-arbitrariness:

---

<sup>28</sup> Many thanks to Christoph Hanisch who pointed out that these two do not necessarily need to be run together and for providing the example.

the status of being constitutive of agency does not suffice for a normatively non-arbitrary status. Of course, if there were some independent reason to be an agent (for instance, rather than a shmagent), or to perform actions, this objection would go away. (Enoch, 2006, p. 43)

Importantly, constitutivists can answer his challenge by appealing to the distinctiveness of the X- feature. Note that this puts them in a *prima facie* stronger position than what Enoch would have us believe. However, it is possible that Enoch could extend his argument to address distinctiveness as well. Imagine I point to a dust ball that has accumulated under my bed and say that rationality norms are constitutive of its distinctive feature. I'm almost certain no one would care to adhere to rationality norms if this is where we ground them.

That is, the constitutivist must locate the source of normativity either in a constitutive feature of agency shown to be independently normativity-conferring; or she must tell us why agency is something we should care about. Constitutivists don't do the former and seem unable to do the latter. And so there are two things that need to come from showing the non-arbitrariness of the constitutive feature: first, a constitutivist would need to show how the constitutive feature he has isolated is unique in that it isn't a constitutive feature of other things such as shmagency. Second, a constitutivist would need to show how it is *this* feature that accounts for the source of rationality norms. But, Enoch argues, in trying to meet these challenges, constitutivists run against the problems of self-vindication (2.5.3) and/ or unavoidability (2.5.4).

### *2.5.3 The Problem of Self-vindication*

A constitutivist might try to show how X is non-arbitrary in the following way: X cannot be arbitrary because when we ask questions about X, we do so by using the very X we are questioning! So, we cannot ask if the features of agency are non-arbitrary if we have to use those features to ask the question in the first place, thus, presupposing the non-arbitrariness of the features. Thus, rendering the status of these features, self-vindicating.



Enoch identifies this as problematic and shows how this is reminiscent of Thomas Nagel's argument that "in order to launch a skeptical attack on logic, the skeptic is going to have to use some logic ...[T]he skeptic's dialectical position is thus rendered unstable" (Enoch, 2006, p. 45). But Enoch does not think that just because the dialectical position is rendered unstable sceptics cannot still pose challenges from this position. He adds that the sceptic is entitled to use logic in virtue of our commitment to the acceptability of its use and that a good *reductio ad absurdum* assumes the correctness of a claim before proving that claim to be false.

Similarly, in the constitutivist's case:

That the skeptic – if there actually is such an interlocutor – cannot avoid engaging the motives and capacities constitutive of agency while mounting his attack only shows that he is in the same boat as the rest of us (surely not a surprise). It does not show that these motives and capacities are self-vindicating or nonarbitrary from a normative point of view. (Enoch, 2006, p. 46)

So, even if it is the case that these constitutive features are employed in asking questions about these features, it doesn't mean that this renders these features non-arbitrary; and their being self-vindicating does nothing to change this.

A possible way to show how these features are rendered non-arbitrary is to show that that they are somehow unavoidable. But, as I explain in the next section, Enoch considers this just as problematic as the claims for self-vindication and non-arbitrariness.

#### *2.5.4 The Unavoidability of Agency*

Although one of the arguably distinguishing marks of agency is autonomy, agency itself is often depicted as an enterprise of perpetual imprisonment. This was already implicit in the last subsection: if we claim that we are unable to frame sceptical questions about features of agency without employing these features, we are claiming unavoidability in the deepest possible sense.

Indeed, for most constitutive accounts of agency, agency is something that happens to us (Korsgaard, 2009; Railton, 2003; Rosati, 2016). Korsgaard claims that we are essentially *condemned* to agency and cannot do otherwise (Korsgaard, 2009). Phrases and commitments such as “we are agents, and that seems to be a very deep fact about us” (Railton, 2003, p. 314), or Enoch who interprets Korsgaard as purporting “the unavoidability of agency” (Enoch, 2006, p. 50); or Enoch’s claim that, “the game of agency [...] is a game you find yourself playing, and one you cannot opt out of” (Enoch, 2006, p. 50). Agency is unavoidable insofar as it is something we find ourselves in and something we cannot escape. This differs from self-vindication in the following way: self-vindication is the role a constitutive feature plays in questioning and answering certain questions. Unavoidability however, is about agency itself – regardless of what you imagine the constitutive feature to be.

How is this related to the *why-care* question? As Enoch argued, one way of grounding the non-arbitrariness of constitutive features of agency in which we source normativity is by showing that we have sufficient reason to be an agent.

So, why should I care to be an agent rather than a shmagent? Velleman (1996) provides an analogy between agency and playing chess.<sup>29</sup> Velleman argues that a person is said to be playing a game of chess insofar as their aim is to check-mate their opponent. And this understanding can be carried over to reasons for acting:

In that case, reasons for acting would be considerations relevant to the constitutive aim of action, just as reasons [...] and anyone who wasn’t susceptible to reasons for acting, because he had no inclination towards the relevant aim, wouldn’t be in a position to act, anyway, and therefore wouldn’t be subject to reasons for acting. (Velleman, 1996, p. 714)

---

<sup>29</sup> The chess example is explored extensively and used (albeit differently from the way I use it here) by Velleman (1996; 2000).

The problem with this reasoning, Enoch urges, is simple. It's all fair and well claiming that a person who plays chess is doing so insofar as they have the right kind of aim, and that a person who is playing the game of agency is doing so as long as they are fulfilling the constitutive aim of action. But something doesn't seem right about this in the sense that a chess player is not unavoidably a chess player, who can never escape his chess-player-ness.

The agency game is simply not analogous to a chess game, Enoch argues. In the chess-playing example, a person could stand up in the middle of the game (perhaps dramatically overturn the board) and exclaim that they no longer wish to play this game of chess. Whereas if the agent takes a knife to his throat and cries out that he no longer wishes to play this game of agency, that act cannot be done outside of agency. An agent can choose to end his life, "but far from opting out of the game of agency altogether, this very decision will be a major move in that game" (Enoch, 2006, p. 50).

Just because agency is unavoidable, however, this does not provide any good reason why we should care to be agents rather than shmagents. Enoch points out that it is entirely possible that an agent, being *condemned* to agency (as Korsgaard so put it), can choose to not play the game *well*. Essentially, our agent can play the game half-heartedly, and "without accepting the aim purportedly constitutive of it as [hers]" (Enoch, 2006, p. 50).

It is possible to go through the motions, because you really have no other choice, without caring one bit about adhering to norms or even caring one bit about being an agent. And rightly, Enoch points out that we need a strong enough account which can show why "the relevant game is worth playing" (Enoch, 2006, p. 52). Enoch believes that constitutive accounts cannot offer an answer and instead, we should just intuitively know that "the fact that our best attempts at

deciding what to do and how to live our lives require normative facts gives us just as good a reason to believe in normative facts” (Enoch, 2006, p. 57).

My response to this is a resounding no. I think it far less likely that just because it would be extraordinarily helpful for me to know how to kill a Parktown Prawn without making a mess and shrieking in terror, that I would somehow be moved to actually *kill* one, or even moved to *believing* that I will be capable of killing one. Because we cannot ground rationality norms here, I soldier on.

So far, we can derive three Enoch-inspired requirements for a satisfying constitutive account of normativity: such an account cannot be grounded either in the self-vindicative character of constitutive features of agency, or in their unavoidability. These are the first three constituents of my Shtest. I now extend the Shtest.

#### 2.5.5 Norm-violation

I think there is an additional criterion, overlooked by Enoch, that is very important to determine if a theory has what it takes to ground the normative grip of rationality norms. In order to develop this criterion, I start with agreeing with Enoch that “our best attempts at deciding what to do and how to live require normative facts” (Enoch, 2006, p. 57). Because this project is aimed at establishing a constitutive account of agency as a source of the normative grip of rationality norms, I maintain that these normative facts can be grounded in what is constitutive of agency. This is to disagree with Enoch’s final claim that normativity cannot be grounded in what is constitutive of agency.

In order to show how normativity can be grounded in any constitutive feature of agency, a first step is to think about how we actually adhere to norms. A few things I would like to discuss

here: the distinctions between norm-adherence and norm-acceptance, and between norm-awareness and norm-acceptance; the relationship between deliberate norm-adherence and deliberate norm-violation; and finally, the relationship between norm-adherence and agency.

### *Norm-adherence and Norm-acceptance*

In Thailand, it is illegal to step on money because the Thai currency has the image of the king imprinted on it and to step on the image of the King is perceived as highly disrespectful. If you mistakenly drop a coin, you best let it roll until it stops naturally. This is considered a norm particular to Thailand and does not apply to many other countries, say South Africa. As a South African, if I were to go to Thailand, I might find it rude that nobody bothered to step on my coin to stop it from rolling into the gutter. Others might consider it rude that I had the expectation that they *should* step on the coin to stop it from rolling into the gutter.

Citizens of Thailand understand that stepping on money is just something they should not do. Perhaps they realised this through noticing that others never step on money or they were disciplined by parents the first time they ever tried to do such a thing. Or, and this is more likely the case, it's just something they don't do in virtue of being born and raised in Thailand. In the first instance, when a child gets disciplined and sternly told that he should not stand on money, he understands that this is wrong and accepts this norm as something he will adhere to in the future. However, in the second case, adhering to a norm in virtue of being something means that no actual acceptance need take place. In other words, I can adhere to a norm without deliberately accepting it as something I should be doing, it is just something I find myself doing.

The distinction between norm-adherence and norm-acceptance is best encapsulated by Kant's notion that there is a difference between acting out of duty and acting merely in accordance

with duty (Kant, 1785). The idea is that acting out of duty is to act out of respect for the norm. So, because not stepping on money is a norm, we adhere to it out of respect of it being a norm (norm-adherence). Acting in accordance with duty however is different in the sense that it is not out of the reverence for the norm that we act but rather another inclination which is unrelated but somehow coincides with the norm. For instance, perhaps I don't like stepping on money because I thoroughly dislike the metallic sound of a coin being stamped on – and that's the reason for me stepping on the coin (lack of norm-awareness).

Sometimes we do the right things without knowing that we are doing the right things. Sometimes we do so purely out of habit. Sometimes we do so in the full knowledge that this is something we should be doing, and we want to do this. Norm-adherence can be all three of these things. Deliberate norm-adherence only happens in the third instance however. When it is coupled with norm-awareness.

*Norm-Awareness vs. Norm-Acceptance and Weak vs. Strong Norm-violation*

In order for deliberate norm-violation to take place, there must at least be some awareness of the norm. It is not necessary that the individual accepts this norm, though she may. In order for someone to adhere to a norm, an individual must accept the norm, and this requires being aware of it. If an individual does not accept a norm but performs actions which seem like norm-adherence, then the individual is acting merely in accordance with the norm rather than deliberately adhering to it.

This is important to this project in the following way: a person could appear to be playing the game of chess by moving the pawns around the table. But unless he is aware of the norms of playing chess, he cannot be accused of deliberately violating chess-playing norms if he moves his bishop straight down the middle of the board. Similarly, he cannot be said to actually be

playing chess at all if he doesn't accept the norms of chess-playing, since if he does not accept them, he can't observe them. For instance, if he doesn't accept that the aim of the game is to checkmate his opponent, then even if his pieces are accidentally in a position to checkmate his opponent, he cannot be said to be checkmating his opponent nor indeed to be playing a game of chess.

Here, I'd also like to introduce two kinds of norm-violation. The first kind is what I would like to call weak norm-violation and the second kind, strong norm-violation. The distinction between the two is important for this discussion. Weak norm-violation refers to norm-violation that is either not purposeful/ intentional or is beyond one's control. Strong norm-violation is the type of norm-violation which is done intentionally and in full awareness of the norm, though I may or may not accept the norm. I would like to introduce four different scenarios to help explain the distinction between norm-awareness and norm-acceptance.

*Scenario A:* I am in Thailand. I have no knowledge that it's a norm in Thailand not to step on money. I drop a coin, and it begins to roll. I don't make a move to step on the coin only because I am wearing roller skates. If I had been wearing shoes, I would have stepped on the coin.

*Scenario B:* I am in Thailand. I have no knowledge that it's a norm in Thailand not to step on money. I drop a coin, and it begins to roll. I am wearing shoes, and I lunge forward and step on it before it rolls into the gutter. People around me are aghast. I am informed by an onlooker that I have done something awful, I have violated a norm – a norm I was not aware of.

*Scenario C:* After the onlooker has informed me that stepping on a coin is violating a norm, I drop another coin in sheer horror that I've been rude. This time though, I want to stand on the coin, but I don't.

*Scenario D:* After not stepping on the second coin, even though I really wanted to, this time I drop yet another coin and I do step on it because I wanted to.

In scenario A, I am not adhering to a norm because I am not *aware* of the norm.<sup>30</sup> Instead I can be said to accidentally act in accordance with the norm.

In scenario B, I have violated a norm but not deliberately so because I had no awareness of it.

In scenario C, I have deliberately adhered to the norm. This in part requires me to have accepted this act of not stepping on the coin as a norm I should follow.

In scenario D, I have deliberately violated the accepted norm.

Scenarios B and D were another pair of examples. In (B), I weakly violated the Thai norm about stepping on money, through not being aware of it in the first place (this of course is different from the last example of texting and crossing at a red light, where I accepted the norm but wasn't aware it applied just then). In scenario (D), I strongly violated the norm.

Now let's introduce one last scenario.

*Scenario E:* Let's say that I have accepted the norm that to step on money is bad. This is something I genuinely care about. I drop a coin. It rolls toward the gutter. I don't step on the coin for one very important reason: I can't, even if I wanted to, because my foot is wedged between two rocks and I am unable to move toward the coin.

---

<sup>30</sup> By awareness I mean as minimal a thing as I can get away with. In particular, I mean either conscious or reflective awareness.



In scenarios C and E, I seem to have adhered to a norm. But this is a mere appearance in scenario E. It's because if I were inclined to violate the norm, I couldn't do it. So then, is it really my choice to observe the norm? The point is this: norm-adherence is not something that an agent *just* does; it must then be as free as her other actions, in order to be genuine. This requires that she be able to not do it – that is to deliberately violate the norm.

What I'm hoping to show here is that in order for genuine norm-adherence to take place, deliberate norm-violation must be possible. I'm going to show why this is relevant in the next sub-section. Deliberate norm-violation and the ability to deliberately violate norms should not be underestimated when we are seeking to establish the normative grip of rationality norms.

Think for a moment of someone who is chained to a traffic light. When the light turns red, she does not cross the road. If the norm is that one should not cross the road when the light is red, we can't say that because she did not cross the road that she adhered to the norm. When she escapes her chains and then does not cross the road, then we can say she adhered to the norm. We can say this because, while she did not cross the road, she could have if she wanted to – thus adhering to the norm, rather than deliberately violating it.<sup>31</sup>

So, what do I mean by deliberate norm-violation? I mean that one is aware of a norm and chooses to act contrary to what that norm requires of one. Scenario D is a good example: I was aware that it is a norm not to step on coins. Yet I stepped on the coin. Deliberate norm-violation requires that I know that what I am doing is wrong, but I freely choose to do so anyway.

---

<sup>31</sup> I appreciate that I am setting a tacit counterfactual condition here that should, in a longer account, be a) more carefully worked out, and b) related to standard versions of the Principle of Alternate Possibilities (starting with Frankfurt, 1969). But I hope that the intuitive point I am trying to make here is clear without getting bogged down in the intricacies of either the vast literature on counterfactuals or on PAP.

### *Deliberate Norm-adherence and Agency*

There are various requirements necessary for deliberate norm-adherence. At a minimum, an agent should be able to form beliefs about the world which are responsive to evidence. Let me begin with an example. An office-worker, who goes by the name of Jay, is asked to form a belief about the very desk he is sitting at. Jay is asked to not just imagine that this object before him is not a desk, but to believe it is not a desk. Jay is asked to believe that this thing before him is, in fact, a car.

This should be very difficult to do for Jay. Some would think that it is impossible psychologically (e.g., Williams, 1973). Others would think it's conceptually impossible: (Railton, 2003; Shah & Velleman, 2005): if the propositional attitude you actually possess towards the object of belief (desk) is this irresponsive to truth-considerations, it is not a belief at all, but something else, like a wish or imagining. But, if it really is a belief you possess, then this belief is incorrect in several ways. Most basically, it is false. But it is also held against the evidence and hence unjustified or irrational.

Typically, for visual beliefs you have no actual *choice* in what it is you believe. You just are that thing which has beliefs which are responsive to evidence (for more on this, see McHugh, 2013; Shah, 2002, 2013; Shah & Velleman, 2005). This may sound unpromising for the possibility of adhering to epistemic norms more generally. But fortunately not all of our beliefs are this automatic.<sup>32</sup>

Let's say that Jay correctly believes that this is a desk. Now, let's say I ask Jay to lift his desk using only a strand of hair to do so. Jay should, as a normal agent, immediately know that this

---

<sup>32</sup> I am deliberately skirting here the doxastic voluntarism-involuntarism debate (e.g., Shah & Velleman, 2005; Steup 2017; Williams 1973).

is an impossible thing to do. Knowing this, if Jay suddenly plucks out a strand of his hair and proceeds to prod at the desk, hoping to lift it, then this is worrisome. This means that Jay is unable to see the link between what it would take to lift his desk and all that is wrong with using a strand of hair to do so. This makes his action irrational. In all ways, Jay is not actually adhering to rationality norms, but he is certainly violating them and violating them in the weak sense.

According to constitutive accounts of agency, if deliberately adhering to rationality norms is constitutive of belief, say, and belief is constitutive of agency, then deliberately adhering to rationality norms is constitutive of agency (Railton, 2003). This I discuss in far more detail in the next chapter. But let's just, for now, intuitively buy into this idea that an agent just is that thing which adheres to rationality norms. If this were true, then how do we explain deliberate norm-violation? Agents are those things which adhere to norms. Norm-adherence presupposes the possibility of norm-violation. But if agents are those things which adhere to norms, and agency is unavoidable, then they can't violate norms. And if they can't violate norms, nor can they adhere to them. This puzzle is quick and nasty, and can probably be resolved by fancier footwork. Nonetheless, it does create a *prima facie* presumption against the constitutivist.

Kathryn Lindeman (2017) argues on independent grounds that the inability to violate norms is a major spanner in the wheel for constitutivists:

any successful constitutivist explanation must provide a metaphysical account of the constitutive features of normatively evaluable kinds that is consistent with Violability. This account must show that there is some connection between the kind and the norms tight enough to account for the constitutive explanation, but weak enough to allow for defect and other evaluations according to norms. (Lindeman, 2017, p. 233)

Transposing this to our context: it is important, in other words, that a successful constitutive theory is able to explain the possibility of norm-violation so that we can be assured of the

possibility of norm-adherence. If an account is unable to explain norm-violation, then it cannot be said to explain norm-adherence.

I will shortly show that existing constitutive accounts (Korsgaard, 2009; Railton, 2003) cannot accommodate strong norm-violation for a very good reason. According to them, an agent is that thing which adheres to norms. If an agent is somehow guilty of norm-violation, especially in the strong sense, then at best her agency is diminished, and at worst, it is absent entirely. This is a false dichotomy: either you adhere to norms and are an agent or you violate norms and are a non-agent. But there is most certainly a third option here – you violate norms and you are still an agent. It should be obvious that setting up a false dichotomy is not a desirable feature of any account of agency. But worse still, if I am right that norm-adherence presupposes the possibility of norm-violation, a theory that sets it up, deprives us of the ability to genuinely adhere to norms at all.

If a constitutive account is only able to explain norm-adherence in terms of agency, then instances of norm-violation (both weak and strong) potentially spell out instances of non-agency. This is counterintuitive as there are many instances, in everyday life, where agents violate norms without undermining agency – such as me knowingly and purposefully stepping on a coin in Thailand. Now of course, this is not the right kind of norm. Social norms are not supposed to be so deeply entangled with our agency in the way constitutivists envisage.<sup>33</sup>

But the same point can be made about morality and rationality norms. Intuitively at least, if I killed someone I wouldn't lose my agency. Nor do I lose it when I do something irrational – say pick up my dead phone to call the phone company to ask them to come and fix it. I say

---

<sup>33</sup> Not unless we held some form of communitarianism about agency such as say Ubuntu, according to which 'A person is a person through other persons' (e.g., Ramose, 2003).

‘intuitively’ deliberately. If you are a constitutivist like Korsgaard, say, you would perhaps think that I do lose my agency in these instances. But this would simply beg the question at hand.

A constitutive account of agency which successfully shows the normative grip of rationality norms, needs to show how norm-violation takes place whilst leaving agency intact. I add this as a fourth criterion that a constitutive account of agency must meet in order to satisfactorily account for the source of rationality norms.

#### *2.5.6 The Shortest Summed Up*

Enoch takes issue with constitutivists who claim we should care in virtue of being agents, and asks: well, why should we care to be agents then? To satisfactorily answer this question, a constitutive account of agency would need to do the following things:

First, it would need to show how the features that are constitutive of agency are indeed non-arbitrary and can confer normativity.

Second, it would need to answer the question of why we should be agents rather than shmagents, without appeal to the idea that these features are self-vindicating.

Third, it would need to show how agency is something we should care about, without appeal to the unavoidability of agency.

Fourth, it would need to show that if we are to consider agents as those things which are governed by norms, then those agents are equally able to violate norms while still remaining agents. Why would it need to do this? Because any account which is trying to show why we

should care to be agents needs to show that, if adhering to norms is constitutive of X (which is constitutive of agency), that these norms really are something we *can* genuinely adhere to.

## 2.6 Objections

### 2.6.1 A Misconstrual of the Unavoidability of Agency

A possible first objection regarding the Shtest is that the problem of unavoidability of agency has been misconstrued. The problem of unavoidability is not tied to *agency*, but rather to the *human condition* and that agency, in and of itself, is considered a solution to the unavoidability of the human condition.<sup>34</sup> Arguably, the challenge of the human condition is what is unavoidable and agency can be viewed as an achievement rather than something which is automatically had. Perhaps, the challenge of the human condition is what Korsgaard alludes to when she claims that we are condemned to action – action being something human beings just do, and one way to deal with this unavoidability is to aim it at something – agency. Two considerations show this concern to be misguided.

First, I don't believe I have misunderstood constitutivists, especially Korsgaard, when she adamantly claims that we are condemned to agency. Being condemned to agency points to agency itself as being unavoidable. Second, if unavoidability is a property of the human condition rather than of agency I have set up a strawman version of constitutivism. But I don't believe I have. I help myself to Enoch's version of what this unavoidability means, and it is the unavoidability of agency. Enoch interprets Korsgaard as talking about agency being unavoidable, and the textual evidence amply supports this view.

---

<sup>34</sup> Many thanks to Christoph Hanisch who pointed this out in feedback.

For example, Korsgaard commits to agency being unavoidable every time she claims that we are condemned to action (the very constitutive feature of agency she argues for) (Korsgaard, 1996; 2009). Railton claims that agency is deep-seated fact about us and something we cannot escape (Railton, 2003). Velleman argues that we cannot even begin to act outside of agency without appealing to the constitutive features of agency to do so (Velleman, 2000). So reading these theorists as thinking that agency is unavoidable is, I believe, justified.

If we buy into the idea that action is constitutive of agency (something I discuss in more detail in chapter 3), then when Korsgaard claims that we are condemned to agency, I interpret this to mean that *agency* is unavoidable – given the constitutive nature between action and agency she has set out. Following Enoch, for the remainder of this project, I treat the unavoidability problem as a problem about the unavoidability of agency – not the unavoidability of something else.

### *2.6.2 Should We Take Enoch's Question Seriously?*

A second group of possible objections is aimed at Enoch's challenge, the challenge I set out to answer in this project. There are two related objections in this group.

The first objection is something we can call the incoherence problem. The second related objection, we can call the circularity charge.<sup>35</sup>

The incoherence problem is this. When Enoch asks why it is we should care to be agents rather than shmagents, he is asking a pre-agent to have reasons to be an agent rather than a shmagent. A worry might be that the question Enoch asks of us is, unfortunately, incoherent. To borrow

---

<sup>35</sup> Many thanks to Jessica Lerm for identifying these problems in feedback provided.

an example from Jessica Lerm<sup>36</sup>, to ask why we should care to be agents is like asking on what day the earth started to turn. There are essentially no days without the earth turning, just as there are essentially no reasons to care without agency. There is something incoherent about asking for reasons to care to be an agent if reasons are the business of agents. This will be an especial problem for me later, when I argue that we have agent-centred reasons to be agents.

My response is three-fold. First, just because there is a mere *possibility* that a question may be incoherent, it does not mean we should not treat it as a serious question (theorists such as Velleman (2010) and Silverstein (2015) have certainly treated the question very seriously). Second, defending the coherence of Enoch's question places an unusually large burden on this thesis and changes the nature of what I intend to do here. I intend to answer Enoch's question, not determine the coherence of the question itself.

Third, Enoch (2011) already has what to my mind is a good response to this challenge. Imagine a paper-sceptical philosopher who writes a philosophical paper arguing that we should not write philosophical papers. As Paakkunainen (2018: 455) notes, "there is a kind of performative self-contradiction in doing so". But, and importantly so, "it's the contents of her arguments that matters, not the fact that she (incoherently) raises them while writing papers" and so, "constitutivists are not off the hook simply by pointing out that the shmagent can't pose that question, or make the request" (Paakkunainen, 2018, p. 456). Therefore, it's not about whether the question can be asked, but rather the content of the request that really matters in this regard.

But maybe this response isn't good enough. Maybe the problem is that it's incoherent to suppose that a non-agent can have reasons, which arguably only agents can have and so the content of the question *does* matter. It's incoherent in the same way that we can't ask why

---

<sup>36</sup> Thank you to Jessica Lerm who provided this interesting analogy to elucidate the problem.



Ostriches don't philosophise really well. It's not in the nature of Ostriches to philosophise well. Axiomatically, they just cannot philosophise, which then makes this question an incoherent one to be asking. I agree that this is an incoherent question to be asking if we can say for certain that non-agents cannot, under any circumstances, be in the business of having reasons just as Ostriches cannot, under any circumstances, be in the business of philosophising. But if there is even a small chance that non-agents can have a semblance of reason-having, then this question is no longer incoherent.

If the reader remains unconvinced with my three considerations, they are welcome to consider my account as one premised on a conditional: if Enoch's question is coherent, then I put forward a substantive account to answer the question. As I have previously stated, to defend the antecedent is another project entirely.

The second problem is a potential circularity charge. When Enoch poses the question: what reasons we have for playing this game of agency over, say shmagency, it appears that he is assuming the existence of a pre-agent. Note, that this pre-agent needs to have a reason to choose to want to play the game of agency over shmagency. This assumes a few things of this pre-agent: a) that this thing can make choices; b) that this thing is capable of having reasons. The problem here, it appears, is that having reasons is, arguably, the business of agency. Note that this objection isn't that the question is incoherent, but simply that it isn't clear what would count as a good answer to it, given that pre-agents can't have reasons or make choices. It's intuitively difficult to think of a creature which is able to have reasons to want to be an agent from outside of a place of agency. Furthermore, it is equally difficult to imagine a creature which can meaningfully make choices etc. without already possessing agency.

To overcome the circularity charge, I argue that choice of agency is not something that only agents can do. If I can show this, I am able to answer the why-care question, I am able to better show what is required in order to ground the source of rationality norms, and I can ward off the circularity charge.

A pre-agent, I imagine, is a subject who has desires and is able to perform actions but is missing the mark of the deliberative nature of agency or missing the mark of fully-fledged self-governance (Bratman, 2007). An example that comes to mind is higher-order animals or children.

According to Priewasser, et al., (2013)

a stable finding in the development of belief–desire psychology is that in some ways young children find it easier to come to grips with desires than with beliefs. Thus, it is sometimes said that young children are “desire psychologists” before they acquire a “desire–belief psychology”. [Furthermore], that folk psychology sees beliefs and desires not just as causes of behavior but also as reasons for acting (Anscombe, 1957, Davidson, 1963). (Priewasser, et al., 2013, p. 546).

We do not, generally, ascribe agency to children, especially young children, because we do not ascribe to them the responsibility and accountability that accompanies agency. But, and importantly for this dissertation, children show that agents are not the only entities capable of reasons to perform actions. Think of a young child, perhaps a six-year-old who actively and knowingly chooses to take on a caretaker role of a younger sibling. Suppose he notices that his parents are negligent, say because they spend their day drinking instead of caring for his two-year old sister. And so the six-year old decides to give the love and care to her that their parents are withholding. Clearly, the six-year-old is not an agent before the decision. We just agreed that we don't treat children as agents. But making that choice is an active buy-in into agency, because it is an active choice of the sort of accountability that characterises agents. Having

desires seems enough for such a choice. A pre-agent, then, can make the choice of agency, contra the objection.

## 2.7 Conclusion

This chapter served to set up the framework and the problems within which a constitutive account which hopes to establish how rationality norms have their normative grip, has to work. Adhering to norms, as well as violating norms, is a quintessentially agential thing to do. Not only is it something agents do, but it is expected that they should be able to do or not do at will.

In order to determine if a constitutive account has what it takes to sustain the normativity of rationality norms, I have set up four problems that need to be overcome in order for a constitutive account to successfully ground the source of normativity. I have grouped these problems under the banner of the Shtest. The first three are borrowed from Enoch (2006).

First, for a constitutive account to successfully ground the source of normativity, it would need to show how the features which are constitutive of agency are non-arbitrary and somehow hold normative-uniqueness. Second, a constitutive account of agency would need to show how, even if the features of agency are self-vindicating, how this is something we should care to choose rather than shmagency. Third, that just because agency is inescapable, that somehow, we can still want to be agents and care to be agents. Fourth, that deliberate norm-violation is as possible as is deliberate norm-adherence.

## Chapter 3 – Constitutive Features: Actions or Beliefs?

### 3.1 Introduction

In the previous chapter, I briefly discussed ways in which we might approach normativity. Most notably, I introduced constitutive accounts of agency. Such accounts aim to establish that standards, aims or motives are constitutive features of agency and then argue that these constitutive features themselves are constituted by adherence to *rationality norms*. I explained that there is one specific problem shared by them all and that is a problem first highlighted by Enoch. In order to ground the source of rationality norms in what is constitutive of agency, constitutivists must show why agents should care to be agents as opposed to shmagents or non-agents. Enoch argues that theorists such as Korsgaard, Velleman and Rosati struggle to address this problem and because of this, cannot successfully ground the source of rationality norms.

In this and the next chapter I examine four constitutive accounts of agency, in order to identify *why* and *where* the failure occurs. This diagnosis will be instructive for articulating my positive account.

In this chapter, I discuss two constitutive accounts:

(§3.2) The first is a constitutive account proposed by Korsgaard which centres around the idea that actions which are governed by rationality norms are actions which constitute an agent. Furthermore, that in virtue of being an agent, these are the only actions available to one.

(§3.3) The second account I consider, proposed by Railton, sees adhering to rationality norms as essential and fundamental to the formation of beliefs; and these beliefs are constitutive of agency.

In section 3.2, I show how Korsgaard's account it does not meet the Shtest criteria. In section 3.3, I do the same for Railton – showing that even though Railton's account misses the mark, it comes closer to it than Korsgaard. Finally, in section 3.4, I argue that because neither of these accounts is in the position to meet the requirements set out in the Shtest, they are unsuccessful at grounding the normativity of rationality norms.

### **3.2 Korsgaard: Self-Constitution as Constitutive of Agency**

One of the most influential and widely referenced constitutive accounts of agency can be found in Korsgaard's works, *The Sources of Normativity* (1996) and *Self-Constitution* (2009). In developing her view, Korsgaard does three important things. First, Korsgaard (1996) proposes that agents are those things which actively engage in 'reflectiveness' and 'endorsement'. Second, there are two principles which an agent cannot *not* endorse. These principles are – adapting from Kant – the hypothetical and the categorical imperative. Third, that actions belonging to agents are immediate endorsements of these principles and are the constitutive features of agency. I'll explain each of these in turn.

Unfortunately, there is no way I can do any real justice to the complexity of these ideas here given space constraints and the vast scholarship on Korsgaard, but I will try to get their spirit right. I should also stress that I do not intend to criticise her view per se, but am rather interested in using some version of it to show three things: i) to elucidate the notion of constitutive connections; ii) to show how there is a possible connection between agency and normativity; and iii) to show how this connection doesn't necessarily hold in light of the Shtest.

At the heart of Korsgaard's theory is the role of action in constituting the self. If action is to play such a huge and foundational role, then we need to understand what exactly Korsgaard means when she talks about 'action'. For one, action is distinct from an act. An act is simply

an event, such as mere behaviour, instinctual reaction, or unintentional behaviour. An action is an act coupled with a reason for acting. Very important here, these reasons are explained by reference to two foundational principles: the categorical imperative and the hypothetical imperative.

According to Korsgaard (Korsgaard, 1996, p. 98) “the categorical imperative [...] tells us to act only on a maxim which we could will to be a law” and “the hypothetical imperative describes what you do when you will an action: you determine yourself to be a cause, the cause of some end. [...] It is a constitutive principle for the will” (Korsgaard, 1997, p. 69).

A well-constituted agent, in this sense, is one who is guided by the constitutive principle. What this means is that without these imperatives at play, what would look like ‘actions’ are in fact, mere acts. Therefore, the constitutive principles are constitutive of the choice of actions. Without these two constitutive principles guiding us, action is no longer considered action. How do we go from act to action? First, we set up ends which adhere to the categorical imperative – the ends being those which we would be happy were it to be universally applicable to all agents. Second, we execute the means to achieve these ends – here adhering to the hypothetical imperative. The idea is that without these imperatives, we would be rendered ‘actionless’. Thus, these imperatives provide us with reasons that constitute our actions.

The coupling of an act and the reason to act has the following consequence: when an agent performs an action, she is *endorsing* these two foundational principles. But from this alone, it is difficult to establish the normative grip of these principles. For instance, perhaps if I am to buy into the idea that as an agent I am endorsing certain principles through my actions, I can choose to endorse other principles instead of these two – which is problematic when trying to explain normative authority – as Korsgaard is wanting to do.

Korsgaard pre-empts this concern and takes a further step. She restricts the volitional capacity of an agent with regards to these two principles alone: an agent is *free* to endorse *other* principles, but just not these two (Korsgaard, 2009). When an agent makes a choice and performs actions, she has automatically endorsed these two principles and has no choice in this matter. So then, any action belonging to an agent is an endorsement of the categorical and hypothetical imperative. The red flag here is that it appears then that an agent is in the strange position to ‘generate’ reasons to justify their actions merely by acting.

For example, by performing an intentional action – regardless of what it is – I can claim to have reason to have performed this action in light of the fact that all actions are dependent on the categorical and hypothetical imperative. So, if asked why I performed an action, my strange (but Korsgaardian) response could be: because I act – since the categorical and hypothetical imperatives are imbedded as automatic reasons for acting. This appears almost tautological and doesn’t seem quite right to explain normative authority.

A possible way to get around this problem is to introduce, as Rosati (2003) does, certain rational capacities which allow for critical reflection on reasons for actions. So, instead of immediately and automatically endorsing the two imperatives through action, at least give an agent room to deliberate and reflect upon the fact that this is what their actions are endorsing. I’m not sure this would be enough, though, without establishing what exactly these rational capacities would look like.<sup>37</sup>

Instead of introducing a reflective and deliberative capacity, Korsgaard offers a potential fix by appeal to constitutivism. It looks like this. If an individual fails to reference the two

---

<sup>37</sup> As briefly discussed in §2.5.2 I am not confident that Rosati herself meets this challenge. But I don’t have the space to defend this claim.

foundational principles in her 'acts', then she fails as an agent for two reasons. First, without reference to these principles, she isn't performing actions belonging to an agent; at most, these things will just be acts. The hope here is to provide a non-reducible basis for actions. Second, she fails at self-constitution.

Why? Because actions which are done by an agent are being caused by the agent in that these are acts which conform to the two foundational (Kantian) principles. These actions are definitive of the integrity of an agent and are definitive of self-constitution, such that the "function of an action is to unify its agent, and so to render him the autonomous and efficacious author of his own movements" (Korsgaard, 2009, p. 161). And so, only actions which are governed by these two foundational principles are actions which constitute agency. She takes this to establish that "the source of normativity [lies] in the human project of self-constitution" (Korsgaard, 2009, p. 4).

What does this kind of self-constitution look like? Korsgaard argues that to say that something belongs to a particular kind is to identify this something with the kind's teleological structure or organisation. It appears then that something is identified as a particular kind by what it *does*. This is how, for instance, we can identify something as a house:

the function of a house is to serve a habitable shelter, and that its parts are walls, roof, chimney, insulation, and so on. Then the form of the house is that the arrangement of those parts that enables it to serve as a habitable shelter. (Korsgaard, 2009, p. 28)

The core notion here is that being a habitable shelter is constitutive of being a house and that "a thing's constitutive function or form *just is* its constitutive norm" (Silverstein, 2016, p. 216, italics in original). When asking why it is that we ought to have a roof on a house, the answer simply is that the roof is essential for the house to function as a habitable shelter.



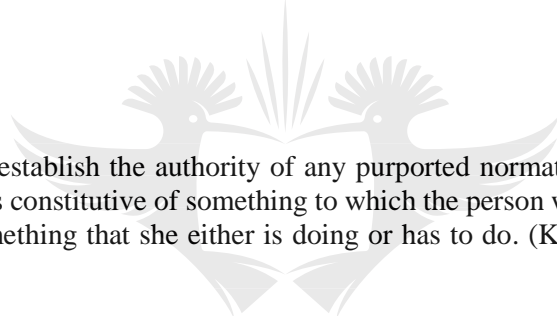
In the same vein, when Korsgaard suggests that action is self-constitutive, she is suggesting that action serves a function of constituting the self and thereby telling us more about what action consists in and what counts as action. Let me consider an interpretation of this claim that I consider most congenial to the purposes of locating the source of rationality norms. Because the constitutive feature in the case of action just is the constitutive norm (Silverstein, 2016), it follows that action serves its constitutive function if it is in accordance with a norm (such as a principle of practical rationality – for Korsgaard – the Kantian categorical and hypothetical imperatives).

If we feel compelled to ask why actions should constitute an agent, the simple answer is that they just do – in the same way that being a habitable shelter is constitutive of a house. I am an agent *by virtue of acting*. Seeing that action is the constitutive function of agency as well as the constitutive norm, it seems clear that we cannot ask the question of why norms govern actions (or why actions should adhere to norms). Actions wouldn't be a constitutive function of agency at all if norms were out of the equation. In fact, these would be mere acts, but would not count as fully-fledged actions.

Korsgaard proposes the constitutive principles – the categorical and hypothetical imperative (Korsgaard, 2009, p. 23). The builder of the house, if not guided by the constitutive principle, would not build a house, rather, a non-house. Although I say a 'non-house' it often sounds like Korsgaard vacillates between two constitutive claims: between X constituting Y and X constituting a *good* Y. Call this the 'good-is conflation', or GIC for short. So, for Korsgaard, if a person isn't guided by the constitutive principle, then she is either not building a house, or she is building a bad house. This has far-reaching consequences for constitutivists, as I will explain later.

For now, let's look at an example of walking. Korsgaard explains that the constitutive principle of walking is that a person places one foot in front of the other. She further claims that unless you are guided by this constitutive principle, you cannot be said to be walking (Korsgaard, 2009, p. 30) So, it doesn't make much sense to ask, of a walker, why they are placing one foot in front of the other. This is just what it is to walk.

If you do not adhere to the constitutive principle, then you cannot be said to be doing that thing. Further, if one were to do something by following the constitutive principle badly, like trying to walk up straight whilst drunk, the activity remains the same, it's badly done, arguably, not being done at all in some instances. (This again, is an instance of the GIC problem.) The crucial idea is that



the only way to establish the authority of any purported normative principle is to establish that it is constitutive of something to which the person whom it governs is committed—something that she either is doing or has to do. (Korsgaard, 2009, p. 32)

This can be construed as the thought that being committed to norms – in the sense that we adhere to norms – is constitutive of action.

To me, the argument looks something like this:

P1<sup>Korsgaard</sup>. Choosing in accordance with rationality norms is constitutive of action.

P2<sup>Korsgaard</sup>. Action is constitutive of agency.

C<sup>Korsgaard</sup>. Choosing in accordance with rationality norms is constitutive of agency.

At face value, this seems like a compelling way of locating the normative authority of rationality norms: we have no choice but to follow rationality norms insofar as we perform

actions which are constitutive of agency. But this kind of all-or-nothing conclusion has its setbacks. This becomes evident when we measure it up to the Shtest. I am now going to go through each Shtest criterion, and see how Korsgaard's account fares. Next, I discuss the four problems as outlined in the Shtest. Again, the main aim here is not to criticise Korsgaard's account, but rather to show that it isn't optimal for grounding rationality norms in agency.

### 3.2.1 *The Unavoidability of Agency*

A potential answer that Korsgaard could give regarding why we should be an agent is that it is something we just find ourselves doing and something we cannot opt out of. But, my intuitive understanding of agency goes something like this: to have agency is to have the capacity to act in a given environment. A person is an agent when they are able to engage with their environment and essentially bring about change through their actions, or in some cases, inaction. Further, a person is an agent when they bring about these changes of their own free choice. It's important that we understand that an agent should have *control* over herself to a large degree – this is what I intuitively understand agency to be about.<sup>38</sup> But if that is so, the idea of agency as something we cannot itself choose is a strange one. Why would the source of choice itself be unchosen?

Support for my misgivings can be found in Enoch (2006), O'Hagan (2014), and Tiffany (2012), amongst others. When we think of agents as being unavoidable and being those things which cannot 'opt out', we are stripping away the very core features of agency – autonomy, control, choice, leaving something behind that looks and acts like an agent (a shmagent), but is in fact, not one. So, we need to modify the definition of agency into something like: an agent is a quasi-

---

<sup>38</sup> And many authors of otherwise divergent views share this intuition, for example, see Bratman, 2007; Frankfurt, 1988.

autonomous individual who is only partially autonomous insofar as she is unable to opt out of this state. Further, that an agent is in control and has free-choice insofar as she cannot stop doing these things; Or, a constitutivist needs to rethink the appeal to unavoidability of agency to answer the why care question. The reason the Shtest sees unavoidability as an issue primarily, is because there is no way to answer the non-arbitrariness challenge if the go-to answer is unavoidability (I unpack this in 3.2.4).

But first, let's see how Korsgaard would deal with the problem of the unavoidability of agency. It should be noted that Korsgaard has a strong stance regarding this: she claims that we are *condemned* to action and choice (Korsgaard, 2009). Korsgaard argues that we cannot *but* act; not acting is just not an option for agents. She argues that even in a situation where we withhold action, that inaction is, in and of itself, still an action insofar as that inaction generates tangible results in an agent's environment and further that this inaction is representative of choice, which is essentially what an agent is condemned to. And so, given Korsgaard's claim that action is constitutive of agency, since action and choice are unavoidable, agency itself is unavoidable. What is the problem here?

The unavoidability of agency is one of the core problems of the Shtest. According to Enoch, to be able to answer the why-care question, we would need good reasons as to why we should be an agent. But it is clear that the agency-unavoidability argument does not, in any way, allow us the room to even ask for reasons to be an agent as opposed to a shmagent. Korsgaard might answer this concern by saying that it really doesn't make sense to ask this question because it is just something that we do. It is nonsensical to ask why we should care to do what we can't help doing in the first place<sup>39</sup>. You are doing it, so deal with it. But this is precisely my concern

---

<sup>39</sup> Important to note that this is reminiscent of the incoherence problem outlined in §2.6. When Korsgaard notes that it is nonsensical to ask this question, she is potentially saying one of two things: first, that it is nonsensical to

(and Enoch's too). What happens if it doesn't matter whether you are an agent in the first place? Perhaps then it won't matter to me if my actions are self-constituting.

### *3.2.2 The Problem of Self-vindication*

So, it is clear that Korsgaard appeals to the unavoidability of agency to answer the problem of why we should care to be an agent, and thus fails to successfully ground the source of rationality norms. Now, for the next problem. On Korsgaard's account, we cannot but adhere to norms of practical rationality when acting. Asking for reasons to follow these norms is an act which cannot but adhere to the very norms it is questioning.

This is, in fact an anti-sceptical tactic used by constitutivists (see also Railton in the next section). If a sceptic asks a constitutivist 'why do you care to be an agent?', the latter will say: 'what a ridiculous question for you to be asking! By asking, you have validated your agency and so rendered the question otiose!'

But as Enoch points out, and I discussed in § 2.5.3, you cannot refute the sceptic in this way. For one option is always to claim that you are assuming the coherence of agency for the purposes of a reductio. And so, because Korsgaard's account essentially stipulates that we are condemned to action and choice, when we ask questions, we are acting and so, through action, we are vindicating agency.

### *3.2.3 The Problem of Non-arbitrariness*

The self-vindication view above is making two important assumptions. First, that because the features of agency are self-vindicating, they are somehow non-arbitrary and unique to agency.

---

ask this question because agents are those things which have reasons and those things which essentially ask for reasons. Second, that to ask is to commit to playing the game of agency anyway – it being unavoidable.

Second, that because these features are non-arbitrary and unique, they imbue norms which are grounded in them with a privileged normative-status. So, because just *these* things are features of agency, *just* these features are such that provide the source of the normativity of rationality norms. But both of these assumptions could be disputed. Unless it made these assumptions, the self-vindication view wouldn't constitute an answer to the question of why we should care to follow rationality norms.

For Korsgaard, the features are very unique and appear, at first glance, to be clearly non-arbitrary. They are essentially self-constituting. Think of the walker again. In order to be said to be walking, you need to adhere to walking-norms. So, you walk in virtue of these particular features (putting one foot in front of the other) and the only way you can do this is by adhering to the walking-norms.

But, one could argue that the same norms which govern walking are also used for running and jumping forward. In each of these instances you are placing one foot in front of the other, but in none of these instances, would you call it walking. This means that the norms which are constitutive of putting one foot in front of the other are not unique to walking. But in that case why would putting one foot in front of the other, be in any special position to imbue *walking-norms* with normativity. But if so, the second assumption made by the self-vindicator is undermined.

Similarly, if the constitutive feature of agency that are supposed to ground rationality norms is something that is shared by other things too, then this feature does not have the non-arbitrariness required to successfully ground the source of rationality norms. Perhaps it is possible that other things – not just agents – also adhere to the constitutive principle and this means that this principle upon which actions are based, are arbitrary.

The onus is not on me to provide empirical evidence that things other than agents adhere to rationality norms when acting or making decisions. However, it's safe to say that higher-order animals such as crows, ravens and chimpanzees have shown considerable skill and deliberation in some decision-making processes, so much so that one could assume a semblance of rationality norms adherence. Almost certainly, especially in instances of sequential tool use to retrieve food, a crow's behaviour seems to adhere to rationality norms (Wimpenny, et al., 2009).

Perhaps machines too (Artificial Intelligence, for instance) can act in a way that adheres to this principle. So, is there a way for Korsgaard to establish these features as non-arbitrary in the sense that they are both unique and have the capacity to confer normativity? The answer is an unfortunate no. To do this, she would need to show that only agents can adhere to this constitutive principle. Given the account she has put forward, this seems unlikely to happen.

Strangely, we may not even be in the position to self-vindicate. If we can accept a world in which what is constitutive of agency is constitutive of other things too, then just because an agent is using the constitutive features of agency to answer the why-care question does not mean they are vindicating agency, they could be vindicating something else instead. For example, perhaps you see a person putting one foot in front of the other and you ask him why he should care to walk. He knows he is placing one foot in front of the other and thinks about what is constitutive of walking (knowing that these things are also constitutive of running and jumping forward) and randomly decides to answer: 'I think I am running'. There is no special hold or grip that walking-norms have over walking if acceptance of the same norms can be seen elsewhere.

#### *3.2.4 The Problem of Deliberate Norm-violation*

If my interpretation of Korsgaard is correct, then it appears that because our actions constitute our agency (P2<sup>Korsgaard</sup>) we are committed, through necessity, to adhere to norms (P1<sup>Korsgaard</sup>), whether we like it or not – leaving little to no room for strong norm-violation – or as a consequence, for deliberate norm-adherence. As I discussed in §2.5.5, the possibility of deliberate norm-violation is necessary for genuine norm-adherence.

In order to understand how important this is to this debate (and to any position which hopes to show that norms are essential to actions or deliberation), we need to imagine that we are requiring something from an individual who could never deliver it and then we further judge him on this fact.

For example, imagine a thief has just snatched a handbag from a defenceless elderly person. The thief decides to take a short cut down an alleyway. At that exact moment, an unsuspecting passer-by stumbles over a rock and bumps into him. The thief drops the bag and continues on his way. Some may herald that passer-by as a hero – someone who retrieved the bag for the elderly person. But could that person have stumbled over the rock, paused for a second and said, ‘wait a minute, I actually want this thief to get away with this crime, let me change the direction of my fall’? Not likely. Why is it that we hesitate to say this person actually did something good? Because he could not have done otherwise.<sup>40</sup>

I suggest that Korsgaard’s account of agency is proposing the same kind of relationship between agency and norm-adherence. Whenever we act, we are adhering to norms, and we just cannot do otherwise. There’s something wrong about calling these actions those of an agent when we are told, that as agents, we have control over the type of things we do – such as bad things or good things. But when it comes right down to it, actually doing bad or good seems to

---

<sup>40</sup> For my disclaimer on the Principle of Alternate Possibilities, see fn. 28.



not be up to us after all, because we never really had any choice in the matter. At least, this is what this particular objection is pointing out.<sup>41, 42</sup>

And again, this is an instance of the GIC problem, the conflation of X being constitutive of Y with X being constitutive of a good Y. It's such that you, as an agent, can't really perform bad actions because if you do then either you're a shoddy agent or you're not an agent at all.<sup>43</sup> Furthermore, if you do something badly enough, you fall out of being that thing in the first place. For example, if the drunk walker walks badly after two beers, then proceeds to drink two more, he begins to stumble about. After six beers, he is lying flat on his face and starts crawling on all fours under the impression that he is walking. Doing something badly enough starts to count for not doing that thing at all.

Let's take a moment to consider the drunk walker who walks badly. Perhaps, if I stretch my imagination, there is leeway here for Korsgaard because in a sense, some norm-violation is happening. But it's not the right kind of norm-violation to count as deliberate norm-violation. The drunk walker is violating norms in virtue of being drunk and thus cannot be said to have full, voluntary control over the violation. It's not up to him, at least in this instance, to pause and say: 'ok, enough drunk walking... time for sober walking'. Deliberate norm-violation is

---

<sup>41</sup> It should be noted that theorists such as Herlinda Pauer-Studer have attempted to show how Korsgaard's account could, with quite a bit of modification, allow for norm-violation or "bad action". Pauer-Studer suggests that we think differently about the categorical imperative. We "can solve the problem of bad action by applying the distinction between constitutive and regulative rules to the categorical imperative" Pauer-Studer. (Pauer-Studer, 2018, p. 37). Fix this reference

<sup>42</sup> Even though Korsgaard has made rather damning claims such as "an unjust person cannot act at all" (Korsgaard, 2009: 152), she does go to great lengths to defuse the objection that her account does not accommodate bad action. For more on bad or defective action in Korsgaard, see Chapter 7, 8, 9 in *Self-Constitution* (2009) where she discusses the notion of a tyrannical soul as "consistently ruled and unified, though it is not self-governed" (2009: 171). Even if she does go to such lengths, I don't believe it's enough to accommodate norm-violation in the way needed.

<sup>43</sup> I am aware that the reader may be concerned that I am setting up Korsgaard's account as a strawman. Arguably, her account allows for degrees of agency where the agent can constitute herself by adhering to the wrong norms. The question then is if the adherence to the wrong norms is deliberate. (Thank you to Christoph Hanisch who pointed this out in feedback.) For Korsgaard, an agent constitutes her agency through adhering to the constitutive principles which is, as it appears, an involuntary agential process. If adhering to norms is an involuntary agential process, then it seems unlikely that the Korsgaardian agent could deliberately violate norms.

the type of norm-violation which should be possible if we are to take norm-adherence seriously. In Korsgaard's constitutive account of agency, someone is an agent only insofar as her actions adhere to norms. This means that an agent cannot voluntarily or deliberately go wrong in respect of the norm.

Like I said before, the constitutive principle of walking is to intentionally and voluntarily place one foot in front of the other. Any violation of this principle, say I flap my arms instead, is not walking. Maybe, Korsgaard would say that the drunk walker is still walking (one foot is being placed in front of the other), it's just bad walking. So, when an agent unintentionally violates a norm, she might say that this agent is still adhering to the norms, but just badly so. When the drunk falls over and is no longer walking, that's when he ceases to be a walker. When an agent stops acting entirely, the agent ceases to exist.

But this seems like a generous interpretation of what Korsgaard is proposing.

But even if the reader doesn't agree with me, my requirement chimes with what Korsgaard herself argues elsewhere. Korsgaard (1997) argues that there can be no such thing as a purely instrumental agent on the grounds that he/ she "cannot violate the instrumental principle. *If we cannot violate it, then it cannot guide us*, and that means that it is not a normative principle" (Korsgaard, 1997, p. 321, my italics).

This is precisely what I have been urging here. I appreciate that she is saying this just about instrumental norms, but why wouldn't the same reasoning apply to any norm? If we accept Korsgaard's claim that a principle can't guide us unless we can violate it, then we surely must accept my requirement that we cannot adhere to a norm unless we can violate it.

To answer this, Korsgaard needs to do one of three things here. First, withdraw the claim that there can be no instrumental agent in virtue of its inability to violate the instrumental principle. Second, she could try to show that an instrumental agent is somehow different from a constitutive agent with regards to the normative grip norms have on either. Third, she could jettison the claim that a constitutive agent can do nothing but adhere to the categorical and hypothetical imperatives.<sup>44</sup> As Lavin points out, “there is a significant difference between the idea of an agent that, as it happens, is not going to go wrong and the idea of an agent that cannot, under any circumstances, go wrong” (2004, p.450).

But perhaps I am wrong about the applicability of this line of thought to constitutive agents. In that case, I will rest content with a more modest and dialectally unproblematic conclusion: Korsgaard accepts the possibility of norm-violation as a requirement on norm-adherence. But her account makes such violation impossible. So, it cannot pass the norm-violation requirement of the Shtest.

As has briefly been touched upon before, Korsgaard attempts to get around this problem by claiming that even though you are constituted by your actions, there are better and worse ways of going about self-constitution, which results in better or worse agents. (See the GIC problem.) This essentially means that an agent is able to violate a norm – at least in the weak sense, but this just means that they’re a poor version of an agent. This, at least to me, is still problematic: either we don’t have a choice to violate norms, or the choice we do have is one that determines whether we are excellent agents or worse agents.

---

<sup>44</sup> Remember, if we violate the constitutive principle, then we cannot be said to be doing that thing. For instance, if we flap our arms instead of placing one foot in front of the other, we can be said to violate the constitutive principle of “walking”. If we violate the constitutive principle of being an agent, by perhaps not acting in a way that is constitutive of being an agent, then we are not agents.

The latter seems even worse than the initial impossible norm-violation scenario I first mentioned. If all my actions are aimed at self-constitution, then by violating a norm am I purposefully aiming at constituting myself as a bad agent? If I am exceedingly drunk and attempt to walk at the same time, I may not know that I am walking badly. In actual fact, I may think that I am walking perfectly well. It may, according to Korsgaard's standards, mean that I am a bad walker, but I'm not sure why I should care that this is so if I'm not even aware of it to begin with. Especially if I believe I am doing perfectly well.

Here's the problem: from an observer's point of view, I'm a shoddy walker and am clearly violating walking-norms. But, from my drunken perspective, I'm doing a fine job at walking. Two things regarding deliberate norm-violation in this instance: to be said to deliberately violate walking-norms, I must, at a minimum be aware that I'm violating the norms. Second, I must also know that this is a norm I don't wish to adhere to – maybe not at this time. In this scenario, I can do neither. First, my perception is that I am adhering to walking-norms – even though I am not. Second, this is a norm I wish to adhere to – I am just currently incapable (and unknowingly) of doing so.

In the previous chapter, I provided the example of a person who visits Thailand where stepping on money is illegal. I discussed various scenarios where a person violates a norm by stepping on money. In one scenario she is unaware that it is a norm; in another, she is aware that it is a norm but chooses to violate it anyway; and in a third, she accepts the norm and adheres to it. One of the morals of this discussion was that in order for someone to adhere to a norm, at a minimum, she must be aware of the norm and accept it. In order to strongly violate a norm, the person must likewise be aware and accept it.

Back to the drunk walker. Two things are going on here. One, he is aware of the norm but is unable to adhere to it (he can't stop himself from stumbling about). Second, he accepts the norm but is not aware that he is violating the norm (he thinks he is walking perfectly fine). If this walker is aware of the norm, accepts it and thinks that he is adhering to it, yet is violating it, is he being a bad walker? From an observer's point of view, the answer is almost certainly yes, because we can see him being a bad walker when he is stumbling about. From his perspective though, certainly not! He thinks he is doing perfectly well and so therefore, in his mind, he is a good walker. The problem is that it may not matter to the walker at all if he is actually a good walker or a bad walker, as long as he thinks he is and as long as he thinks he is adhering to these norms that govern his walking.

Two things here. First, should an agent not have some kind of acceptance of the constitutive principle (the categorical and hypothetical principles) for him to fully endorse these through his actions? Second, does it matter, in the least, to the agent that his actions are self-constituting? Minimally, there needs to be some kind of buy-in in order for real endorsement to occur. Korsgaard pushes the point that when an agent performs an action, it is in virtue of endorsing the constitutive principle. So, when we see agents act, we see norm-endorsement. When a walker walks, we see walking-norm endorsement. When a drunk walker stumbles about, it's a display of non-walking-norms endorsement.

Korsgaard may respond that she doesn't care about acceptance or even awareness. Rather, if you're performing actions, then you've already endorsed those principles and there's no way you could do otherwise. But then this completely removes the possibility of deliberate norm-violation. Because of this, I think self-constitution in this sense is automatic and not deliberate. If anything, I am inclined to think it leaves the agent impotent and ineffectual in the story of self-constitution. Exactly what Korsgaard is not aiming for.

In light of my concerns above, my suspicion is that she cannot defend her position on self-constitution, at least, given other commitments that she has. But since I promised not to do Korsgaard scholarship, let us suppose she can make either of the replies above. Unfortunately, it leaves the agent in an even worse situation: this person is not a walker at all and isn't aware at all that he isn't one.

Let me explain. Korsgaard seems to be committed to something like the following argument:

P1<sup>Korsgaard</sup>. Putting one foot in front of another is the norm for, and hence constitutive of, walking.

P2<sup>Korsgaard</sup>. Walking is constitutive of being a walker.

C<sup>Korsgaard</sup>. So, putting one foot in front of another is constitutive of being a walker.

If this is right, regardless of which of the two replies above Korsgaard gives, it is not the case that the drunk is walking badly. He is not walking at all. On the first possible reply, he isn't adhering to norms by stipulation; on the second possible reply he isn't adhering to walking-norms, because he isn't aware of them or doesn't endorse them – both requisites for adherence. But if he isn't adhering to such norms, then - by P1 – he isn't even putting one foot in front of another. By P2, he isn't walking either.

Two things to note about this result. First, it is implausible: drunk people – who are still standing more or less – do put one foot in front of another, just wobblingly; and do walk, just badly. Second, this result illuminates the sense in which Korsgaard's account rules out the possibility of norm-violation: as long as you are a walker, you put one foot in front of the other, but as long as you do that, you adhere to walking-norms.

Let's say that I have the required awareness to wilfully constitute myself as a good agent, rather than a bad one, then I wonder why I would care to constitute myself as a good agent rather than a bad one. There must be some kind of value in being a good agent that surpasses the value of being a bad agent for me to want to be one rather than the other. Put another way, suppose an agent is convinced that Korsgaard is right and that the aim of all of his actions is self-constitution. We can still ask, with Enoch

Do we have any reason to believe that now he [the agent] will care about the immorality or irrationality of his actions? Why isn't he entitled to respond along the following lines:

Classify my bodily movements and indeed me as you like. Perhaps I cannot be classified as an agent without aiming to constitute myself. But why should I be an agent? Perhaps I can't act without aiming at self-constitution, but why should I act?. (Enoch, 2006, p. 41)

Considering Korsgaard's account, through norm-violation, an agent does one of three things:

1) She ceases to perform 'actions' which belong to an agent; 2) she establishes herself as a shoddy agent; 3) Ceases to be an agent. We are met with a few severe ultimatums here: Either you adhere to norms and perform actions which are agent-worthy or those things which you are doing are not actions at all; Either you adhere to norms and are a good agent or you violate norms and you're a bad agent; Either you adhere to norms and you are an agent<sup>45</sup> (or a lesser agent) or you violate norms and you are a non-agent.

But in any case I hope to have shown here that it isn't even clear that these are the right terms of the ultimatums. For if we can't violate norms without losing agency, then we cannot adhere to norms either. As Korsgaard herself concedes, norm-adherence presupposes the possibility of violation.

---

<sup>45</sup> Phillipa Foot (1985) and Judith Jarvis Thomson (2008) have developed similar versions of constitutivist accounts to that of Korsgaard with varying degrees of success regarding this problem I discuss here.

### 3.2.5 Korsgaard has failed the Shtest

One of the most contentious questions in discussions about how norms have authority over our actions and agency is why we should even bother or care about norms in the first place. It's not just that norms *have* this peculiar authority, but it's difficult to show *how* they have *any* authority. Yet until we have shown this, it remains mysterious why we appeal to norms in the formation of our beliefs, in our deliberations and decision-making and in our actions etc.

If we are to accept a Korsgaard-style account, then we just must take norms seriously based on the following consideration: “the fact that we are at risk of falling apart if we become bad enough is supposed to show that those norms have authoritative force for us; we cannot help but care about them” (Lindeman, 2017, p. 247). Supposedly, we should be afraid of becoming such terribly shoddy agents that we somehow fall out of agency. Arguably, this seems like a legitimate concern (see §3.2.1). But as I have argued, with Enoch, the problem goes deeper than just being shoddy agents. Korsgaard's theory needs to make it clear why we should care to be agents at all. Given the various responses I imagine Korsgaard could give, it is evident that she falls against all four problems set out in the Shtest: the problem of deliberate norm-violation, the unavoidability of agency, the problem of self-vindication, and finally, the problem of non-arbitrariness.

I don't see a way out for her account (even in the best-case scenarios I have already proposed) and I believe (and would like to argue for such) that any constitutive account that derives its constitutive feature from action will face the same problem. It isn't clear that there is a satisfactory way to unite action with norms in a way that will answer these problems – at least not in the way Korsgaard approaches it.



But perhaps the problem isn't with constitutive accounts per se, but rather with the particular feature that Korsgaard has chosen. So, let me consider a different view – one that posits *belief* as the constitutive feature. Hopefully it will fare better.

### 3.3 Railton: Belief as Constitutive of Agency

Railton's account is this: aiming at truth partly constitutes belief; and belief is partially constitutive of agency.<sup>46</sup> Railton proposes this view as a way of providing a non-hypothetical account of why persons must appeal to practical or theoretical reasoning to guide actions and deliberations. Such an account is supposed to provide us with non-hypothetical reasons for caring about norms, reasons that is, that don't depend on the particular contingent goals individual agents might have. A constitutive account neatly fits this bill, as it grounds such norms in essential and universal features of agency.

Railton presents this constitutive account in the following way. First, he shows that belief is a propositional attitude which “cannot represent itself as unresponsive to – unaccountable to – [...] truth” (Railton, 2003, p. 297). The constitutive nature of belief is such that propositional attitudes which are not aimed at truth – or at least which are not responsive to admitted evidence – are not beliefs. Persons who are unwilling to accept this ‘price of admission’ to belief, for instance, someone who forms beliefs based on random bits of information and who is not committed to truth – would not be in the position to believe the propositions at all. Therefore, according to Railton, “the special relation between belief and truth thus comes with the territory of belief” (Railton, 2003, p. 297).

---

<sup>46</sup> Views like Railton's, which presuppose that belief aims at truth (e.g., Williams 1973), are to be distinguished from the more recent view that knowledge can be understood as the aim of belief (e.g., Williamson, 2000; Simion, et al., 2016). I use Railton's because he makes explicit the constitutive connections between belief and agency on the one hand, and between belief and norms on the other. But nothing I say here commits me to a view on what the actual aim of belief is.

Railton's next step is to argue that the process of forming beliefs is guided by norms. For example, (not Railton's), imagine you believed, for a considerable part of your life, that Austria was a country famous for its favourite animal, the kangaroo. You proceed to get into a heated discussion with someone about strange-looking animals and you mention the strange kangaroo from Austria.

The person you are conversing with immediately corrects you and points out that you are thinking of Australia. Not Austria. Further, the person provides proof by showing you pictures of kangaroos in front of the Sydney Opera House. You do an internet search and find that you have been confusing Austria and Australia! You are obligated, to forego the previous belief you had and adopt this new one. To retain both would be to have two contradictory beliefs. Contradictory beliefs cannot both be true. So, if you are to keep them in full consciousness, you'd be undermining their status as beliefs. But, of course, not having contradictory beliefs is one of the central rationality norms. So, to have beliefs – attitudes aimed at truth – comes for free with adhering to rationality norms.

Second, Railton shows that beliefs are constitutive of agency: (2003, p. 298) “an agent acts on intentions and plans, which constitutively involve beliefs and are formed deliberately in part on the basis of beliefs”. So, beliefs are necessary for intentions, since you cannot have intentions without belief. Since intention is a core feature of agency, the conclusion is that without beliefs, one cannot be an agent.

If aiming at truth is to be understood as adhering to rationality norms, then Railton's overall argument goes like this:

P1<sup>Railton</sup>. Adhering to rationality norms is constitutive of belief.

P2<sup>Railton</sup>. Belief is constitutive of agency.

C<sup>Railton</sup>. Adhering to rationality norms is constitutive of agency.

I appreciate Railton's argument because I believe it can be interpreted as showing a strong connection between normativity and agency. It differs from Korsgaard's in the following important way: where Korsgaard's agent is one which endorses a constitutive principle through action, for Railton, we follow rationality norms insofar as we form beliefs which are constitutive of agency. Where Korsgaard's account struggled to address the problems outlined in the Shtest, I think Railton's account fares better because of the nature of the constitutive feature he proposes.

### 3.3.1 *The Shtest*

The Shtest, recall, demands of a constitutive account of agency that it provides us with sufficient reason as to why we *should care* to be agents rather than shmagents. To answer the why-care question, a constitutivist account must address the following problems: it must show how norm-violation is possible; it must not appeal to the claim that agency is unavoidable to show why we should care; it must not appeal to the self-vindicating character of the constitutive features of agency in explaining the feature's non-arbitrariness or norm-uniqueness. Let's see how well Railton does.

### 3.3.2 *The Problem of Norm-violation*

While Railton, too, faces the problem of deliberate norm-violation, it is to a much lesser degree than Korsgaard. Since they are both constitutive accounts, it would seem tempting to assume that both should equally suffer the objection that they do not allow for deliberate norm-

violation and as such, cannot explain what would appear as irrational action or belief. But Railton's account stands ready against this objection in the following ways.

First, he admits that any account which does not allow for the possibility of norm-violation has the "unintended effect of pulling the claws of the very criticism" (Railton, 2003, p. 69) it is supposed to ground. That is, an agent who fails to have attitudes which are aimed at truth, fails to have beliefs (P1<sup>Railton</sup>), and hence ceases to be subject to (epistemic) rationality norms, since such norms are meant to govern beliefs. Thus, such a person cannot then be a target for rationality-judgements/ assessments.

The way in which Railton overcomes this problem is to claim that we cannot deny someone agency just because they fail to completely satisfy this constitutive feature, labelling this as "xenophobic" (Railton, 2003, p. 71). This is to call for a less rigid constitutive accounts of both agency and belief: both "come in degrees, and are not all or nothing" (Railton, 2003, p. 71).<sup>47</sup> He recognises too, however, that we cannot afford to be too flexible in a constitutive account either, or else we risk not being able to pinpoint where the normative grip of norms lies.

It appears then, that a certain degree of belief is constitutive of agency. Anything less than that degree of belief and agency falls away. For Railton, then, the challenge is finding just the right kind of constitutive relationship between belief and agency – and by doing so, finding a way to allow for deliberate norm-violation without compromising the existence of agency. But the question is: does he succeed?

For Railton (2003), one could ask whether "we have reasons for agency versus non-agency as a way of being" (Railton, 2003, p. 75), and concluding the discussion by simply pointing out

---

<sup>47</sup> The quote here just concerns belief. But it is clear from the discussion that he thinks of agency as a graded affair, too.

that if we are led down that path by asking if “*being an agent* is the best or only way of getting what [one] most wants from life” (Railton, 2003, p. 75), we encounter an extraordinary consequence, because by asking what is best for your own life, you are essentially engaging in this deliberation as an agent. Therefore, it appears that one cannot even begin fathoming non-agency for oneself from the perspective of an agent. That would be an impossible pursuit – in fact, one would be hard-pressed to show how this kind of deliberation would *not* appeal to norms, and so it appears incoherent to ask why we should care about norms. I will return to this point below (§3.3.4), as it has obvious bearing on the self-vindication element of the Shtest.

Being too rigid in a constitutive account threatens to show too much and being too flexible results in no constitutive account at all. Railton aims for just the right mix. One way of getting this is to note the distinction between aiming at truth and being true. What is constitutive of belief, of course, is the former. What this does is allow for deliberate norm-violation in the sense that belief aimed at truth can still fail to be true, and fail because a norm has been violated.

There are three distinctions to be drawn here. First, beliefs are attitudes aimed at truth, but this does not necessarily mean these beliefs are always true. Second, there are truth-norms, such that one ought to only believe truths. Third, sometimes rationality norms miss their mark. Importantly, these three things float free from one another. A failure in one does not automatically mean a failure in another. Therefore, it is possible to violate one but still adhere to the others.

Imagine that the first time I ever saw a picture of a kangaroo was in a book where the author misspelt ‘Australia’ and wrote ‘Austria’ instead. I never cared enough about kangaroos to check any further sources to confirm this information. My having the belief that kangaroos come from Austria is false, even though the attitude itself is aimed at truth. By holding an

incorrect belief, I am violating a norm. Katsafanas explains this as “what constitutes an attitude as a belief is aiming at truth. But what constitutes an attitude as a good belief is achieving that aim, by being true” (Katsafanas, 2013, p. 62).

Again, I need to talk about the two different kinds of norm-violation here. To say that belief aims at truth and misses its target is one kind of norm-violation. When I believe that kangaroos live in Austria on the basis of textual evidence, I was adhering to truth-norms – I did not purposefully seek out this falsity and adopt this false belief in the face of contradictory evidence. I had the correct attitude of aiming at truth – I just happened to miss the mark by believing the wrong thing.

Unfortunately, though, this is still a norm-violation. It just seems to be weak norm-violation rather than strong norm-violation. It is not as though I am aiming to break the norm of truth; I just inadvertently do. I can't, in fact, deliberately violate this norm while retaining the attitude of belief, because to do so would be to undermine belief's aim.<sup>48</sup> For imagine, after being shown evidence, over and over again that kangaroos are in fact from Australia, I bluntly refuse to change my belief accordingly – that would be deliberate norm-violation.

A strong kind of norm-violation would be possible with a different set of norms – norms that tell us how to live up to belief's chief norm. Here we can think of following the evidence or avoiding contradictions. And here it seems that deliberate norm-violation *is* possible. Thus, I can deliberately avoid the evidence for certain beliefs – e.g., that my partner is unfaithful – by not going to certain places, avoiding certain friends, etc.

---

<sup>48</sup> On one reading, anyhow. There are many interpretations of what it means for belief to aim at truth. For a comprehensive collection on the topic, see Chan, 2013.

When I formed the belief that kangaroos originated from Austria, it wasn't my intention to form a false belief. Hence, I didn't violate the truth-norm deliberately. When I am presented with new information, that I have reason to trust, I change my beliefs accordingly. Strong norm-violation would happen the moment I encounter evidence that kangaroos do not originate from Austria, but I like my long-held belief better and do not change my belief – even though I know I should. The problem with this is that if my beliefs become generally unresponsive to truth, then those attitudes are actually not beliefs at all, but something else instead. We are back with Railton's attempt to get just the right mix. If we can accept that sometimes the aim of belief does miss its mark and this is where norm-violation arises, then it appears that Railton comes close to overcoming the problem of norm-violation.

Unfortunately, though, his theory does not do as well with regards to the other problems outlined in the Shtest.

### 3.3.3 *Unavoidability of Agency*

So far, I've explained how Railton's constitutive account shows a flexible relationship between belief and agency which aims at alleviating the worry about norm-violation. This means that Railton's account fares better than Korsgaard's. But, I will now argue, it still misses its target: when it comes to answering why we should care to be agents, rather than something else, Railton, like others, appeals to the unavoidability of agency.

He imagines an eternal sceptic, Gary. When Gary asks why he should want to be an agent rather than a non-agent, Railton responds that it seems a rather strange question to be asking since “after all, we *are* agents and that seems to be a very deep fact about us. Justification has to start somewhere, and if it is to be justification *for us* it had better start where we are” (Railton,

2003, p. 74). From within the position of agency, we can ask for justification, but not from outside of agency – not in the way Gary is hoping for.

Even if we consider a scenario where an agent is forced into a position to elect euthanasia rather than to live with a terminal disease, we need to consider if this choice is “to put an end to his agency” (Railton, 2003, p. 74). Railton argues that the choice itself is done within the confines of agency. “When [Gary] asks, in effect, whether the exit price is worth paying, he is asking whether being an agent is the best or only way of getting what he wants most from life” (Railton, 2003, p. 75). But this deliberating about the exit price can only be done from the perspective of considering opting out of one’s current agency-status – there’s just no way to look at agency from an observer perspective and say something like: ‘that thing over there, I don’t want it’ (pointing to agency); ‘but this thing over here, I want it’ (pointing to shmagency).

Here, we can see that Railton is committing whole-heartedly to the unavoidability of agency and hence falling prey to the associated problem. At least, unlike Korsgaard, Railton provides reasonable explanation as to why it is we experience this unavoidability of agency. So, instead of stating that we are condemned to action (as Korsgaard proposes), Railton argues that

the pleasures (or other advantages) on the other side of the borderland couldn’t possibly count for you. They wouldn’t be yours. You’re an agent – that’s one of the deepest facts about you –and they’d be the pleasures of a non-agent. (Railton, 2003, p. 75)

So, even *considering* the life of a non-agent seems ludicrous as it wouldn’t matter to you, as an agent. This leads to the third and fourth problems of the Shtest.

### 3.3.4 *The Problem of Self-vindication*

When Gary (Railton’s sceptic) asks why we must be agents rather than slipping into non-agency and Railton explains that it is because whatever is happening outside of agency can



never be relevant to you. The response has one further consequence. When Gary asks his questions, he is using a particular means/ends form of reasoning, which Railton identifies as the Low Brow sort. According to this view, practical reasons are considerations that promote the desires of the individual.<sup>49</sup> By asking why one should not just abandon agency and become a non-agent, Gary is asking for practical reasons to be something other than what he is. But by asking this question, he is conceding to having desires which need to be sated. As Railton has claimed, it doesn't make sense to ask if non-agency is a better enterprise given that the desires of the non-agent would not be yours, so effectively you cannot ask questions about whether it would be *better* for you to be something other than an agent.

Conversely, even though Railton does not run this argument, we could run a similar argument for the High Brow view. By asking for reasons to be an agent rather than a shmagent you are asking what is so good about being an agent. Under both the Low Brow and the High Brow view, we are left without sufficient answers.

Railton shows how deeply rooted beliefs are in agency. First, agents are those things which intentionally act, often by endorsing an end and taking the means to accomplish this. Furthermore, that our agency spanning over time depends on beliefs about our memories and expectations. All of these things, Railton argues, “constitutively involve beliefs and are formed deliberately in part on the basis of beliefs” (Railton, 2003, p. 58). To imagine a world where we replace these beliefs with something else, such as shbeliefs maybe, is to effectively eradicate that which makes up the anatomy of our agency. There is no way for Gary to show how shbeliefs can do the job of beliefs without completing foregoing agency. And so, “paying the price of admission to belief is necessary to gain entry to agency” (Railton, 2003, p. 59) .

---

<sup>49</sup> The High Brow view holds that all practical reasoning is aimed at the good. See my discussion of the guise of the good in §2.4.3.

Gary cannot effectively ask for practical reasons outside of agency to give us reason to be agents – practical reasons are not going to do this for us.

Two things come from this. First, you cannot ask for reasons to be an agent outside of agency or from the perspective of a non-agent – because by asking this question, you already concede the point of being an agent. Second, obviously your desires matter to you because you are caring if the world as a non-agent would be better for you – meaning you either want more than what you have now or less than what you have now – but in either instance, you are asking if non-agency is better based on what is currently going on with you as an agent. So, you cannot doubt the force or point of practical reason without engaging in the very thing whose point you are purporting to doubt. Unfortunately, it appears then that Railton cannot disentangle himself from the problem of self-vindication.

### 3.3.5 *The Problem of Non-arbitrariness*

What essentially makes this a better account than Korsgaard's with regards to the Shtest is that it appears that there is a uniqueness to the constitutive features of agency in Railton's account, which is not seen in Korsgaard's. Remember, we are looking to establish how the constitutive features of agency are non-arbitrary and further, that these features can confer normativity on rationality norms. Railton comes close, at least with the former.

First, by showing that your *desires* as an agent are an important thing for you to consider when deliberating about whether to be a non-agent. Second, that as a non-agent, you will not be in the position to enjoy those things which only an agent can enjoy. In this way, Railton has shown that desires of an agent is a better thing to possess than the "desires" of a non-agent. This does, in some sense, establish that these desires are non-arbitrary, but I don't think in the way required by the Shtest.

The Shtest requires that we show how these desires are such that they are unique to agency, and that they are imbued with a normatively-privileged status, in a way other desires are not. However, Railton does not identify desires as the constitutive feature of agency. He identifies belief as this feature instead. What is needed here is to show that beliefs are in the position to inform desires, rendering them non-arbitrary.

For example, it is the case that I may share similar – very arbitrary desires- with my cat. Perhaps, we both desire to nap and have regular meals. But just because we share the same kind of desires, I cannot say that my cat has agency. What is the difference between my desires and my cat’s desires? One way to show the difference is to argue that I, as an agent, have desires which are impacted or informed differently from what impacts or informs my cat’s desires. Arguably, what informs my cat’s desires is instinctual, biological needs. What informs my desires may very well be beliefs.

This places Railton straight up against the self-vindication problem – which further pushes him against the problem of non-arbitrariness (I explain this in a moment). When Gary asks about why we should be an agent, he does so from the perspective of means/end reasoning, and Railton responds to him as such:

you already defer, in posing this question, to the very thing you seek to challenge. You must already see – and feel – the ‘practical logic’ of what you claim to find arbitrary or problematic: the bearing of ends upon means. (Railton, 2003, p. 77)

It appears then that Railton can only establish the non-arbitrariness of the constitutive feature of agency, by appealing to self-vindication. This is problematic, because we should not be establishing non-arbitrariness in virtue of self-vindication. To get around this, Railton would either need to establish non-arbitrariness and somehow forego self-vindication or commit to

self-vindication and somehow forego non-arbitrariness. But, because of the intimate connection he has established between the two, this seems an impossible task.

Railton's account cannot show non-arbitrariness without appealing to both the unavoidability of agency and self-vindication. By arguing that adhering to rationality norms is constitutive of beliefs and that beliefs are constitutive of agency, Railton establishes norm-uniqueness and normativity conferral insofar as the constitutive feature is non-arbitrary. But remember, this is something I have argued to be problematic. Because a certain feature is constitutive of agency, it does not mean it holds some kind of norm-uniqueness status, nor does it mean it has normativity-conferring powers. As Enoch argues, there is a lot more needed to establish this norm-uniqueness or non-arbitrariness. For one, we need to be able to show why I should care to be an agent. If I can do this, then I come closer to showing how that which is constitutive of agency somehow holds a norm-uniqueness non-arbitrary status. Railton is unable to do this because of the intimate link he forges between self-vindication and non-arbitrariness.

Overall, then, Railton has a better grip on some of the issues than Korsgaard does, but can he convincingly tell us why we should care to be agents in the first place; and finally, why should one accept rationality norms and hold beliefs rather than shbeliefs? Enoch notes that: "perhaps I cannot opt out of the game of agency, but I can certainly play it half-heartedly, indeed under protest" (Enoch, 2006, p. 188). I'm not sure what Railton's response would be to this specifically, but I imagine he would say that we should care to always aim at having true beliefs with the intention of always hitting the target, because that's just what it means to be an agent – and that's just what it means to have beliefs. If we don't do these things, we run the risk of slipping into the realm of non-agency – which would mean nothing to us because the experience of the non-agent would not count as experience for us.

I believe Railton's account is better able to deal with certain requirements set out in the Shtest. For instance, it is able to accommodate norm-violation far better than Korsgaard's. However, it still isn't able to withstand the other challenges put forward by the Shtest. So then again, maybe this particular constitutive feature – belief – isn't right either. In the next chapter, I discuss a different kind of constitutive feature: a drive.

### **3.4 Conclusion**

In this chapter, I examined two prominent constitutive accounts of agency and measured the rate of success of each according to the Shtest.

First, I looked at the constitutive account of agency proposed by Korsgaard (§3.2) and concluded that Korsgaard fails the Shtest. Korsgaard holds that agents are those things whose actions are guided by the categorical and hypothetical imperatives. She is unable to address the norm-violation problem without simultaneously either diminishing agency or sacrificing it entirely. In addition, Korsgaard encounters the problem of the unavailability of agency when answering the why-care question and she commits more strongly to this than Railton does by claiming that we are condemned to action and essentially have no way to opt out of agency. And because of this notion of unavailability, we ask the why-care question from the place of agency, thus facing the challenge of self-vindication. Finally, Korsgaard is unable to show how these constitutive features of agency allow for norm-uniqueness, but instead says that they are normatively-unique in virtue of their constitutive nature. This, I argue, we cannot buy into.

Second, I examined Railton's account (§3.3), which proposes that belief is constitutive of agency. Railton deals with the problem of deliberate norm-violation a bit better than Korsgaard by not committing to the diminishment or sacrifice of agency, but instead maintains that when deliberate norm-violation takes place it is because the aim of belief essentially misses its target.

Further, Railton commits very strongly to the notion of the unavoidability of agency, but unlike Korsgaard provides a better explanation of this unavoidability – maintaining that any benefits accrued on the side of non-agency would not matter to us, at least providing some reason to care to be an agent.

Because Railton thinks that by asking for reasons to be an agent, you accept the value of practical reason and hence agency, Railton also commits to self-vindication. Railton also fares a bit better than Korsgaard with regards to showing how these constitutive features hold norm-uniqueness status. By identifying how important desires and benefits are to an agent as opposed to a shmagent and by linking this to the notion of beliefs, it seems to be the case that beliefs, which are constitutive of agency, are indeed normatively-unique, at least more so than is the case with Korsgaard's account.

In the next chapter, I discuss two further constitutive accounts of agency. I argue that they do much better at the Shtest than either account examined here in chapter 3. But even so, they still miss the mark.



# Chapter 4 – Constitutive Features: Drives?

## 4.1 Introduction

The primary constitutive accounts of agency I discuss in this chapter centre around the idea that there is a unique drive which is the hallmark, constitutive feature of agency. Velleman's account claims that having a drive for sense-making<sup>50</sup> (Velleman, 2000; 2013) is constitutive of agency. At the end of this chapter, I briefly discuss a second account (rooted in the continental tradition) proposed by Paul Katsafanas (2013). Katsafanas argues that having a drive for power is constitutive of agency.

This chapter is comprised of six sections. Section 4.2 will briefly recap the four demands that the Shtest makes of any constitutive account of agency which hopes to ground the source of rationality norms. Section 4.3 will show how Velleman's account gains ground on the others, in meeting the problems indicated in the Shtest. In section 4.4, I provide a diagnosis. I first show how each of the constitutive accounts discussed have either fared better or worse than the other. I then diagnose the critical flaw shared by all accounts, which explains why none can adequately meet the Shtest-challenges. In section 4.5, I briefly examine yet another account, from a different tradition (continental philosophy), proposed by Katsafanas. Finally, in Section 4.6, I conclude the chapter.

## 4.2 The Shtest and its Demands

---

<sup>50</sup> I use Mitova's (2016) term 'drive for sense-making' throughout this project. Other ways to refer to this concept is: drive for intelligibility, or drive to make sense.

The burden of Enoch's (2006) attack is this: a constitutive account of agency will succeed in grounding the normativity of rationality norms only if it can answer the question of *why we should care to be an agent*. This forms the basis of the Shtest.

As I have argued, to answer this question successfully, a constitutive account must be able to do the following:

*i. Show how norm-violation is possible;*

*ii. Do not appeal to the unavoidability of agency to answer the why-care question;*

*iii. Show how the features which are constitutive of agency are indeed non-arbitrary;*

*iv. Do not appeal to self-vindication to answer the why-care question.*

In the next section, I examine elements of Velleman's account to see if he passes the Shtest. I conclude that Velleman gets closer than the previously discussed accounts, but still misses the mark.

#### **4.3 Velleman: The Drive for Sense-making**

Velleman offers a constitutive account of agency which is quite different from both Korsgaard's and Railton's, in that he situates the constitutive bit in a primary, non-reducible drive or unique desire. This non-reducible, *sui generis* drive is to make sense of what one is doing and to do only what makes sense. He develops this idea in the following way. Firstly, he highlights that what is wrong with the standard model of action is that it leaves the *agent* out of the story.



The standard story of human action, according to Velleman (1992b) is that an agent identifies something she wishes to have or attain, she further identifies actions which will get her this thing, and then she proceeds to perform these actions accordingly. The problem with this account, for Velleman, is “that the story fails to include an agent” (Velleman, 1992b, p. 461). The account makes it look as though human action *happens* to the agent rather than the agent bringing about the action.

A way around this problem is to make use of a hierarchal model of action, such as the one proposed by Harry Frankfurt (1971; 1988). Hierarchal views hold that there are different levels of desires or motivations which bring about action. Frankfurt identifies first and second-order desires. A first-order desire is a desire or wish to perform such and such action. A second order one is a desire to have or not to have some first-order desire (Frankfurt, 1971:7).

Velleman argues that the hierarchal model of action also leaves the agent out of the story. The problem here is this:

UNIVERSITY  
OF  
LONDON

It doesn't matter, in the hierarchal model, whether the subject is satisfied with his first-order motives because of depression or boredom or laziness, or, alternatively, because it is responding to their force of reasons. (Velleman, 2000, p. 14)

Again, the agent is missing from her own actions. It is not clear how endorsement of a first-order desire is supposed to be representing the agent any more than the first-order desire itself that is. Finally, to remedy the problem of the ‘missing agent’ in the story of action and accommodate the idea that an agent is that which authors their own actions, Velleman proposes a constitutive account of agency in which action has a constitutive aim, modelled on the way belief has a constitutive aim (truth). (See Railton in chapter 3.)

For Velleman, the constitutive aim of action is *the aim of knowing what we're doing* and doing only what makes sense. To elaborate, if Frankenstein was tasked with infusing his monster

with agency, he would have needed to design his monster “to gravitate towards knowing what they’re doing, and they will only do those things which they have made up their minds that they’re going to do, and so they will act by choice” (Velleman, 2000, p. 26). Essentially, the action which is most *intelligible* for the agent to take will be the one that she is typically driven to choose and so “action is thus behaviour aimed at intelligibility” (Velleman, 2013, p. 133).

But, what of actions which are surprising to the agent? Say, for instance, that I am informed that a close family friend has been imprisoned, and upon hearing the news, I burst out laughing.<sup>51</sup> This action takes me by surprise.<sup>52</sup>

Perhaps I never realised until now, how much I blame this family friend for being the reason for my parents’ divorce, for example. And so the laughter is a response to this feeling I’ve had all along, without ever fully realising I harboured such feelings. It makes sense to me to burst out laughing – although this malicious action might not be understood by observers. That action is authored by me. This is full-blooded action.

Some may argue that making sense requires, at a minimum, some kind of acceptance of rationality norms (Mitova, 2016; Wedgwood, 2007). Think for a moment if the agent did not accept rationality norms, instead she just did things based on whatever whim caught her fancy in the moment. Or, in the decision-making process, she was guided by a Magic 8 Ball. She would simply be in no position to make sense of what she is doing.

---

<sup>51</sup> Some may argue that what I have explained as a surprising action is actually not a proper action because it isn’t done for a reason, at least not in the traditional sense. According to Hursthouse, this kind of ‘action’ is such “that the agent would not have done the action if she had not been in the grip of whatever emotion it was, and the mere fact that she was in its grip explains the action as much as anything else does” (Hursthouse, 1991, p. 59).

<sup>52</sup> I am aware that laughing is not a standard example of an action. However, I use it here for simplicity sake to point out that there instances where we may know something and it comes as a surprise that this is something we do rather than what would be considered a more appropriate action/ reaction.

For example, perhaps she asks the Magic 8 Ball if she should marry the next person who walks through the door and the Magic 8 Ball throws back the following answer: ‘it is decidedly so’. The reason for marrying the next person who walked through the door is contingent on one of twenty possible responses a Magic 8 Ball can offer, which are completely random. Perhaps, you could say that she did have some authorship in the action of marrying – she *asked* the Magic 8 Ball and followed the recommendation. But this is not the type of authorship being pushed by Velleman. It seems unlikely that you can rationally justify making such a huge life decision on the answer a toy throws at you.

I suggest that sense-making requires *consistency and reliability*. For example, at a minimum, for an agent to make sense of what she is doing, she requires beliefs which are responsive to evidence and are not contradictory with each other (Mitova, 2016).<sup>53</sup> These are two norms of theoretical rationality. If an agent did hold contradictory beliefs, then an action would not make sense to her.

Mitova (2016) argues first, that the drive for sense making according to Velleman is a hallmark of agency. Second, to have the drive for sense-making requires us to form beliefs about what makes sense for us to do. Conversely, if I weren’t able to form beliefs, then I would not be able to make sense of what I am doing – hence, my actions would fail to be intelligible to myself. So, Motiva points out that the formation of beliefs itself must be guided by epistemic norms in

---

<sup>53</sup> Mitova (2016) argues that the drive for sense-making requires, at minimum, that it is guided by epistemic norms in order for it to work the way Velleman imagines it would. My argument here is similar except for the idea that what is also required for the drive for sense making to do the work it is set out to do, a further requirement is that of reliability.

order for those beliefs to perform their function.<sup>54</sup> Beliefs that are not guided by epistemic norms are, arguably, not beliefs at all.<sup>55</sup>

Velleman argues that autonomy is ‘conscious control over one’s behaviour’— that is being conscious of one’s behaviour and of being in control of one’s behaviour –Mitova claims that, “the drive to sense-making is constitutive of conscious control” (Mitova, 2016, p. 212) such that if you were to perform an action instinctively, like catching a knocked-over glass, the action is only truly autonomous if you had enough control over the movement to stop yourself catching the glass if that is what you willed to do (Velleman, 2000).

Anything that does not manifest such control is not considered fully autonomous action. Mitova points out that “I have this type of control only when what I am doing makes sense to me” (Mitova, 2016, p. 212). And so, such control is a manifestation of the drive for sense-making. Without the drive for sense-making, we would not have the type of control necessary to be considered full blown actions belonging to those of an agent.

I agree with Mitova that “such beliefs making sense just amounts to their being the product of following things like the evidence- and consistency norms” (Mitova, 2016, p. 213). However, I argue that it is not enough that beliefs are responsive to truth and do not contradict one another – thus satisfying the consistency criterion. A further requirement, I argue, is reliability.<sup>56</sup>

---

<sup>54</sup> I acknowledge that this is a very condensed version of Mitova’s argument and does not do it justice. But for the sake of this project, I believe it is enough to show that there is a constitutive link between epistemic norms and that which is the constitutive feature of agency.

<sup>55</sup> For more on the role of epistemic norms in belief formation, see Cote-Bouchard, 2016; Littlejohn, 2012; Railton, 2003; Toppinen, 2015.

<sup>56</sup> For lack of space, I cannot go into much detail regarding reliabilism. For more on reliabilist epistemology, please see: Dretske, 1970; 1981; Dunn, 2015; Goldman, 1986; 2008. For a prominent internalist critique, see, e.g., Feldman 1985. If readers are unhappy about the externalist ring to the proposal I am making, they are welcome to fall back on the point about evidence and consistency. This point does not depend on taking sides in this debate.

When I speak of reliability, I mean:

An object (a process, method, system, or what have you) is reliable if and only if (1) it is a sort of thing that tends to produce beliefs, and (2) the proportion of true beliefs among the beliefs it produces meets some threshold, or criterion, value. (Goldman, 1986, p. 26)

Richard Feldman (2003), although an opponent to reliabilism, provides to my mind a great example of why reliability is attractive when forming beliefs. He tells the story of two bird watchers taking a stroll in the woods. One is a novice bird watcher and the other is an expert. They both see, at the same time, a pink-spotted flycatcher. Both form a belief that they are seeing a pink-spotted flycatcher. With regards to justification, the expert is justified in forming this belief because of her years of experience – her bird-perceptual apparatus is reliable. The novice, by contrast, jumps to the conclusion because he is excited and happens to guess correctly (Goldman & Beddor, 2016).

Methods of forming beliefs can be either reliable or unreliable. The expert's process of belief-formation is reliable because she is able to match the visual appearance of the pink-spotted flycatcher with images she has remembered seeing in the past – perhaps in textbooks. The novice forms the belief based on an unreliable guess.

For Mitova, sense-making beliefs are guided by epistemic norms such as evidence and consistency norms. But, let's say that the novice bird watcher goes out into the woods a second and third time, perhaps even ten times, and he correctly guesses a whole variety of birds (again, from pure excitement), he is effectively succeeding in being guided by evidence and consistency norms. He is responding correctly to evidence and is consistently identifying the correct birds. But this is not a reliable way to form beliefs. I argue that the novice bird watcher is not being completely intelligible to herself even if he adheres to the epistemic norms Mitova

proposes. To be fully intelligible, the birdwatcher would also need a reliable way to form these beliefs instead of just reacting to evidence and not being inconsistent in this belief.

If sense-making beliefs are formed as a result of following epistemic norms such as being responsive to evidence, being consistent, and being reliable, then “our inventing [rationality norms] consists in our framing a conception of them that has epistemic authority” (Velleman, 2004, p. 228) and further that “a person’s conception of what he is doing has epistemic authority because he tends to behave in accordance with it” (Velleman, 2004, p. 229). Velleman draws attention to G.E.M Anscombe’s distinction between two thoughts one might have about the future: “I am going to be sick” versus “I am going to take a walk” (Velleman, 2013, p. 129). Velleman points out that the former is not a decision at all, but rather a prediction of something that is going to happen that is beyond the agent’s control. The latter is a decision which is very much in the agent’s control. For Velleman:

This contrast is analogous to that between reading and writing: ‘I am going to be sick’ is a *reading* of the speaker’s future, whereas ‘I am going to take a walk’ is a case of *writing* it. That’s why we think of ourselves as the authors of our actions but not of other behaviour. (Velleman, 2013, p. 130)

Importantly, when this agent thinks that they are going to take a walk, they do so because that is what makes sense for him to do – thus, aiming at intelligibility. Intelligibility, as Velleman argues, hinges on the correctness of beliefs, such that if “a subject who lacked a disposition to follow indications of truth would not be capable of forming beliefs” (Velleman, 2013, p. 133). And as I have argued, there are certain rationality norms that need to be adhered to in order for the drive for intelligibility to be manifest in the sense Velleman proposes.

And so, if aiming at sense-making is to be understood as constitutive of agency, then the way Velleman’s account grounds the normative authority of rationality norms goes something like this:

P1<sup>Velleman</sup>. The drive for sense-making is constitutive of agency.

P2<sup>Velleman</sup>. Adhering to rationality norms is constitutive of the drive for sense-making.

C<sup>Velleman</sup>. Adhering to rationality norms is constitutive of agency.

I now consider how this account fares against the Shtest.

#### *4.3.1 The Problem of Norm-violation*

If a constitutive account does not allow for norm-violation, then we cannot say that a person genuinely adheres to norms. I argued that some constitutivists (Korsgaard and Railton) failed to meet this requirement by entailing that if you violate norms, your agency is either diminished or lost.

For Velleman, an agent is that which strives to act in a way that makes sense. He provides the following example: a child pushes a glass off of a table and you instinctively catch it compared to a cheekier child who tosses the glass to you and says ‘catch’! Essentially the behaviour will be the same – in both instances you extend your hand to catch something. The difference between the two ‘actions’ is that the former is unintentional, and the latter would be “a fully intentional action” (Velleman, 2000, p. 189). In this example, in the first instance, the agent is not performing an action which is intelligible to herself. There’s no chance for sense-making there. Because the formation of our beliefs is dependent on following norms – such as consistency, evidence, and reliability norms (as I have argued in the above section)– it appears that the drive for sense making comes with adhering to these norms.

So then, what does violating rationality norms look like under this account? According to Mitova (2016), just because we sometimes experience a lapse in sense-making, it does not

follow that agency is undermined, since we can still be aiming for sense-making. The drive for sense-making may be unsatisfied but is still present. It is possible that when norm-violation takes place, it is because the agent either did not make sense because she did not fully accept rationality norms (although she was aware of them as rationality norms), or that the agent had a lapse in sense-making, where the norm-violation occurred non-purposefully.

I think Velleman's account fares far better than Railton and Korsgaard's regarding the problem of norm-violation: we can violate norms while still aiming to make sense, and hence agency is not undermined. Velleman asks us to

keep in mind that self-understanding is not simply a matter of making sense of oneself as one is; it is also a matter of making sense to oneself, perhaps by being otherwise (Velleman, 2013, p. 32)

[and] the fact that someone cannot find it in himself to do something does not entail that it isn't there to be found, that no reframing of his self-understanding would make the act intelligible for him. (Velleman, 2013, p. 111)

Velleman's account explains weak norm-violation (norms that are mistakenly violated) very nicely. Think of a pedestrian crossing the street while reading a text. If she happened to make it across the street alive, she might experience a moment where she looks up and realises the light is red. She may pause and ask of herself: 'why did I just do that?' – thus, trying to make sense of an action she obviously had little control over. But this does not mean her agency has fallen away.

The example above points to weak norm-violation. Arguably, strong norm-violation can take place, too. I imagine a scenario where the pedestrian is texting and is walking, being aware that this is a bad thing to be doing given that there is intersection coming up where she may be required to stop if the light should be red. She continues to text even in light of this evidence – texting is what makes more sense in that moment than to worry about the light turning red.



The norm being violated in the above example is a legal norm as well as a rationality norm. The legal norm is that crossing an intersection when the light is red is against the law and would be frowned upon by any onlooker. The rationality norm is that there is enough evidence to form a belief that this is a very dangerous exercise, but she chooses to do it anyway. Velleman (Velleman, 2013, p. 108) maintains that “a person may feel categorically forbidden to do something because it is, as we say, unthinkable” and so too, unintelligible. However, it is also possible that one can reframe the unthinkable so that it makes more sense for the agent to do.

For the texter, perhaps she sees the red light and knows that it is forbidden and unthinkable to cross the intersection whilst the light is red. However, she is also texting a parent who has been hospitalised and the hospital happens to be within walking distance – across from an intersection. And so, she reframes the unintelligible scenario in such a way that it becomes more intelligible for her to not care about the red light or the intersection she is about the cross, as she is in a hurry to get to the hospital.

#### 4.3.2 *The Unavoidability of Agency*

The Shtest expects that the constitutivist will not use the unavoidability of agency to answer the why-care question. This means that the constitutivist can't say that we are agents and so, because of this, we just do care. Let's see if Velleman is able to stay away from this problem. Velleman does commit to certain aspects of agency being unavoidable, specifically with regards to autonomy and self-understanding. Frist, he explains autonomy as being a “part of your personality from which you truly cannot dissociate yourself” (Velleman, 2000, p. 30) and also, your understanding is your “inescapable self, and so its contribution to producing your behaviour is, inescapably, your contribution” (Velleman, 2000, p. 31). Furthermore, “you are

obliged – condemned, if it comes to that – to make up your future course of action” (Velleman, 2013, p. 131).

Two key elements in Velleman’s constitutive account of agency is the idea that autonomy is our ability to purposefully act in a way that makes sense to us. This requires intimate understanding of ourselves and what makes sense in the context of our goals and values. That is why Velleman’s account is to do what makes sense for oneself, doing what is most intelligible to do – resulting in self-knowledge. He proposes that autonomy and understanding are key constitutive features of agency, and that these things are inescapable, or that from which you cannot dissociate yourself. He argues that we should care to adhere to norms because norm-adherence is constitutive of our drive to make sense of what we are doing; and this drive is inescapable. Caring is just something that we do as it is in the constitutive make-up of what we are (Velleman, 2013, pp. 130 - 138).

If this is so, then why should we *care* to be agents in the fully-fledged way in which Velleman pushes for in his constitutive account? Enoch (2011) specifically challenges Velleman on this by stating that “it’s not generally true that engaging in some activity – satisfying some relevant descriptive criteria – suffices for having reason to direct oneself at its constitutive aim” (Enoch, 2011: 211). Velleman answers that

the aim of self-understanding is inescapable for you, and in two senses. First, it is naturally inescapable for you as a human being [and] because the aim of self-understanding is naturally inescapable, the question whether to have it is moot: you cannot help having this aim, like it or not. (Velleman, 2013, pp. 136 - 137)

Therefore, you are a creature which aims for sense-making; if you were to play the game half-heartedly, you’d be aiming to half make sense; but this is meaningless and just not something an agent does – at least not given the constitutive aim of the drive for sense-making. So not fully playing the game of agency is not an option – you are either in or out; but you can’t be

out. But this still does not mean that just because we unavoidably have the aim for sense-making we should care to have it, or that we should have reason to have this aim.

Given the close link Velleman has established between sense-making and action, it also just seems unlikely that we *could* even play the game half-heartedly to begin with. Let me explain.

Velleman claims that

when we choose or decide what we are going to do, we settle that question in our minds and we thereby settle the same question in fact. We make up our minds that we are going to do something, and we thereby make it the case that we're going to do it... [we] will be inhibited from doing anything until [we] think [we] are going to do it, and then [we] will be prompted to do what [we] think. (Velleman, 2000, p. 22)

What this means is that an individual has a reason for performing an action and this reason is a fact “that would provide an integrative knowledge of what we are doing, if we did it” (Velleman, 2004, p. 233). To see what he means, think for a moment of our deliberations as being almost predictive in nature. To use Velleman’s terms, we write our futures as opposed to reading them (see the ‘going to take a walk’ example in § 4.3). Reasons for action are closely tied to what it makes sense to do. So, to answer the why-care question, Velleman writes:

When the agent thinks about what it would make sense for him to do in light of his circumstances, attitudes, and attributes, he cannot honestly purport to be reading his future in them, since what he does is going to depend on what he sees as making sense in light of them. [...] The very thoughts by which the agent might try to read his future are thus thoughts by which he cannot help but write it. (Velleman, 2013, p. 131)

By asking the why-care question, one is exemplifying the very hallmark of agency Velleman proposes. It is, in this sense, unavoidable. I don’t know if this would be sufficient an answer for Enoch, because I think Velleman is falling into the trap of the unavoidability of agency here – just perhaps not as hard as Korsgaard and Railton do. I say this, because instead of the answer Korsgaard gives: you care just because you are an agent, and Railton: you care because being

an agent is a deep fact about you; Velleman is saying: you care because you have a reason to act in a way which is essential for agency. You are the author of your future actions in light of what makes most sense for you to do.

Perhaps Enoch might respond, ‘Alright, so somehow you do what is most intelligible for you to do and somehow, this drive for sense-making is guided by norms. But why should you care about this drive? Why not care about a shdrive instead?’ Velleman’s response to Enoch is clear, but still, according to Enoch, unsuccessful:

‘why be agents rather than shmagents?’ Are you contemplating agency and shmagency as alternatives between which you are trying to make an autonomous choice? Or are you viewing them as alternoids between which you are trying to make an autonomish shmoice? If the former, then you are an agent, and the criterion for your choice is set by the aim constitutive of your agency; if the latter, then you are a shmagents, and the criterion for your shmoice is fixed by the shaim constitutive of your shmagency. (Velleman, 2013: 143).

For Velleman, if you are trying to make a choice and considering alternatives, then you are thinking from within the narrative of agency. However, if you are trying to make a shmoice and considering alternoids, then you are thinking from within the narrative of shmagency. Therefore, making a choice to choose agency is already done from within a place of agency.<sup>57</sup>

Enoch responds to this by pointing out that Velleman is possibly treating the question of why care to be an agent in two different ways: first, that agency is naturally inescapable – which Enoch concedes (2011: 216), but also claims is not important in this context; second, that agency is dialectically inescapable.

Let’s deal with the first way of thinking about the why-care question. Enoch concedes that agency may be naturally inescapable, but believes that even if this were the case, it does not

---

<sup>57</sup> Note though, that I do not accept this claim in §2.6.2.

follow that just because a person *is* one thing, that somehow, and automatically so, we care to be that thing. But worse still, even if one were to care to be an agent, it doesn't then follow that one has reason to be one. He provides an example of a "latent and grudging patriot" who rejects all forms of patriotism but still finds himself caring about patriotic things (like national teams or the national anthem) (Enoch, 2011: 216). The idea here is that the patriot cares about things that are patriotic but this does not give him a reason to care about those things.

But perhaps the more interesting inescapability is the dialectical inescapability Velleman pushes for. Dialectical inescapability is such that there is no viable standpoint from which an agent could ever launch the inquiry into finding reasons to be agents – there is no place from which an agent can have access to agent-independent norms or agent-independent choices. This is very closely linked to self-vindication and this is why, I argue in chapter 5, this type of inescapability allows for the problem of self-vindication to hold. According to Velleman, to ask why we should aim to make sense is already an episode of making sense (Velleman, 2013: 137).

Enoch answers by making use of an example of a paper-sceptical philosopher (already discussed in §2.6). The paper-sceptical philosopher is determined to get the message out there that philosophers should stop writing papers but should write books instead. She believes that philosophers, when writing papers, often do so with a superficiality that wouldn't be present if philosophers wrote books instead. However, ironically so, the paper-sceptic philosopher proceeds to publish her findings in a paper. Enoch claims that we cannot now discard all her hard work and the content of the paper just because it is written as a paper – essentially challenging the medium she used and ignoring what it is she expressed in the paper. And so

The analogy, I hope, is clear. Showing that the practical-reason-sceptic (the one asking 'Why should I care about (e.g) self-understanding) has no safe grounds from

which to launch his attack is neither here nor there. It does not even begin to vindicate practical reason. Thinking otherwise is like settling – in the discussion with the paper-sceptic – or noting that she’s written a paper, without tackling her arguments against paper-writing head on. (Enoch, 2011: 220).

Thus, calling on the unavoidability of agency to answer the why-care question is not a good move. Unfortunately, this is something Velleman wholeheartedly commits to.

#### *4.3.3 Non-arbitrariness*

To avoid the problem of non-arbitrariness, Velleman would need to show how the drive for sense-making is non-arbitrary, and further, how this drive has normative-uniqueness status, unlike, say, other desires. I have already suggested that Velleman clearly shows the relationship between practical rationality and the drive for sense-making. However, establishing the normative-uniqueness status of this drive for sense-making needs a bit more work to pass the Shtest.

Velleman proposes a scenario where we are designers of a creature with the ultimate goal of imbuing this creature with agency (see the earlier Frankenstein example). He asks what it is that will be the final tipping point from being a mere creature to being a full-blown agent. He first introduces practical reason into the creatures and claims that you “would add it in the form of a mechanism modifying the motivational forces already at work. You would design practical reason to survey a creature’s motives” (Velleman, 2000, p. 12). Furthermore, he identifies this mechanism as that which will distinguish a nonautonomous creature from an autonomous creature, that is as the tipping point between a mere creature and an agent. Most importantly however,

considerations of these motives and circumstances are what qualify, in my view, as reasons for acting. They are the considerations out of which we can fashion a description that would embody a knowledge of what we are doing, if we applied that description to ourselves in the way that would prompt us to behave accordingly.

Reasons provide us with an account of what we could be doing, and, indeed, *would* be doing if we adopted an expectation to that effect. (Velleman, 2000, p. 27)

Velleman asks of us what we would include in the creation of an agent and concludes, along with practical reasoning, that we would need to include a special kind of desire to make sense of the world. So, this agent has the desires and beliefs as any other creature, but also a special desire which motivates it to act by driving it to make sense of the world and ensure intelligibility. This is the distinguishing mark of an agent as opposed to a non-agent. The good thing about all of this is that this special desire (or drive) is universally shared by all agents because to be an agent is to be constituted by this drive – which means, as Velleman imagines, we can ground normativity.

The drive for sense-making is understood as the drive which makes autonomy possible. The desire for sense-making is a second order one, which is in a position to arbitrate amongst the first-order ones. Velleman imagines

the drive toward self-understanding as exerting a fairly minor, modulating role in our practical affairs. It influences which desires or objects we choose to pursue, how we harmonize them with one another, organise our efforts toward them, and express our thoughts and feelings along the way. (Velleman, 2013, p. 28)

At the heart of Velleman's account is the following central claim: making sense of what one is doing, is that which ensures conscious control over one's actions, and in turn, a grasp of agency. But does Velleman's account deliver all it says it does? When Velleman creates his agent, he inserts this drive into the agent and magically the agent pushes through life, making sense of herself and her surroundings, and in the act of trying to figure out why she is doing these things, she is relying on her drive to ask these questions. But why, drawing on Enoch's argument here, should this created agent actually care to make sense at all, rather than non-sense, or shmense? One possible answer is that the value of making sense of things in the world is that we are

better able to succeed in our tasks (both small and big). But this is the type of instrumental value we are trying to avoid – we are looking for a constitutive one.

The idea is that there is some kind of motivational drive which enables him to seek out that which makes most sense for him to do – providing a reason for an action. The problem with this however is that it isn't clear how, just because something makes most sense to do, that it is the thing we should do. Furthermore, that if two things being equal make most sense to do in any given situation, which one is best? Is there a way, in Velleman's theory, to rank which reason is best? I explain this below.

As much as Velleman promotes the drive for sense-making as having a 'modulating role' of first-order desires, it isn't clear how this modulation happens. For Velleman, desires upon which we act,

appear to provide the strongest reason for acting, then the desire to act in accordance with reasons becomes a motive to act on that desire, the desire's motivational influence is consequently reinforced. (Velleman, 1992b, p. 479)

First, if we view the drive for sense-making in this light, it runs the risk of falling against the problems associated with neo-Humean views (see §2.4.2). Second, a clear problem with viewing the drive for sense-making as that which mediates between the first-order desires is that "Velleman has no account of how some reasons for acting can be stronger than others" (Magill, 1997, p. 163).

And if so, the problem of non-arbitrariness rears its ugly head. I think, to show how the constitutive feature of agency is truly non-arbitrary, at least in the sense that it can pass the Shtest, we need to show how the constitutive feature of agency is somehow able to "provide us with a way of ranking reasons" (Katsafanas, 2013, p. 74). Why? Because Velleman's



theory is incapable of generating conclusions about what we have more reason to do. For the mere fact that action constitutively aims at self-knowledge does not imply that we have more reason to act in ways that generate more self-knowledge. (Katsafanas, 2013, p. 75)

And so, if Enoch is right in his line of questioning, then a constitutivist would need to show why, and even how it is that we have reason to be an agent rather than a shmagent. If we have reason to be an agent and reason to be shmagent, which reason is weightier?

If a constitutivist can provide us with this answer, then I believe non-arbitrariness can be established. But if she has no way of ranking reasons then the project is doomed from the beginning.

An interesting consequence of this line of questioning, is that the constitutive could turn the tables on Enoch. If Enoch is demanding reasons to be an agent rather than a shmagent, then the constitutive could very well retaliate with the same demand: provide for us, Enoch, reason to be a shmagent rather than an agent. It becomes quite clear, that if we were to buy into the ranking of reasons to be one thing over another, that we may very well have stronger reasons to be an agent than to be a shmagent. But this can only be done if the constitutivist has a way of ranking reasons. If not, the demand is meaningless when articulated by her.

#### *4.3.4 The Problem of Self-vindication*

Why should we care to be agents? For Velleman, simply put, by asking the question of why care, we are moved by the drive for sense-making – for without the drive, we would not feel compelled to have asked this question in the first place:

to ask ‘why should I have the aim of making sense?’ is to reveal that you already have it. If you do not seek to do what makes sense, then you are not in the business of practical reasoning, and so you cannot demand reasons for acting or aiming. (Velleman, 2013, p.137)

Thus, Velleman's account is self-vindicating in the deepest possible sense. In the very act of asking why it is we should have the constitutive aim in the first place, we are making use of that very constitutive aim to do the asking (Velleman, 2013, p. 137).

For example: if the game of playing chess, is to checkmate one's opponent, then as long as a person has this aim, they are playing chess. In the same way that if the game of agency for instance, is to aim to make sense of what one is doing, then as long as a person aims to make sense of what one is doing, then they are playing the game of agency. Part and parcel of playing the game of agency is to ask questions about why it is that we do one thing rather than another. Asking why we should care to be agents rather than shmagents is to already play the game of reason-giving, and hence to accept the need to be an agent as well as the normative authority of reasons. This is a perfect example of self-vindication.

What if asked whether we can opt out of the game of agency? The answer is (as discussed at length – especially with regards to Korsgaard and Railton) that simply through asking if we can opt out, we are doing so from the place of agency – thus, again, establishing that agency is self-vindicating. But there is also something rather peculiar about asking if we can opt out. By asking why we should continue, this intuitively seems to be a question about choice. Consider the following: should I continue playing this game of agency *or not*? The '*or not*' implies that we have a choice (this I explore in far more detail in the next chapter).

Because Velleman argues that agents are those creatures which write their futures, rather than read their futures, he is committing to the idea that we automatically have a reason to be an agent rather than a non-agent (although the ranking of reasons still remains a problem here). And so, by answering the question of why care, one may be asking why we care to continue in the game of agency rather than opt out. For Velleman

When I form a choice or decision, however, I aim to settle a question in my mind insofar as I can thereby settle it in fact; I aim, that is, to avoid representing an arrangement that I am not thereby managing to make. In the face of compelling evidence that I'm not going to do something, my mind cannot remain made up that I'm going to do it: the decision to do something, like the belief that I'm going to, cannot withstand evidence to the contrary. It thus represents things as having been arranged in some way, with the aim of representing how they really have been arranged, albeit by this very representation. (Velleman, 2000, p. 25)

When we ask why we should care to not opt out of agency, we are asking why it is that we should not just stop what we are doing and perhaps do something else instead. What I get from the above quote is that the *decision* is what makes most sense to do. I mentioned in §4.3.2, that our deliberations are almost predictive in nature, in the sense that if I desire to do something and I establish that I shall do it and that it is something which makes sense for me to do, then it is something I will act out.

It appears then that I have a conception of what my agency looks like and this conception may feed into my actions. Also, given the predictive nature of our thoughts – for instance, ‘I am going to take a walk’ feeds into the reason for me to take a walk – then having a certain conception of what my agency looks like, might feed into the reason for me to be an agent.

Velleman writes that

a rational agent tends to enact the attitudes and traits that he conceives himself to have, by pursuing what he thinks that he wants, through means in which he thinks that he believes, and in ways characteristic of other dispositions that he ascribes to himself. (Velleman, 2013, p. 90)

Hence, when an agent answers the why-care to be an agent question, she already has an answer which is settled in fact – it is predictive in nature. By asking why I should care to be an agent, I am asking this question in a way that is self-vindicating. First, I am asking because I am trying to make sense of my agency. I can only be making sense if I have the drive for sense-making, which is the constitutive feature of agency, according to Velleman. Second, I already have a conception in my mind of what agency looks like for me and I cannot entertain another

conception – such as entertaining the notion of shmagency – which is contradictory to my first conception.

Consider this passage:

if by thinking ‘I am going to take a walk’ you can make it so – and if you can also make it not so by thinking its negation – then your thoughts are not answerable to any fact of the matter, even if there is one. Let the fact of the matter be that you are going to take a walk: that fact must be partly due to the prior fact that you are not going to do otherwise; for if you were going to think otherwise, then you would subsequently do otherwise. (Velleman, 2013, pp. 130 - 131)

Therefore, Velleman’s (Velleman, 2013, p. 137) response to why we should care to be agents is: “to ask ‘why should I have the aim of sense making?’ is to reveal that you already have it”. This is a blatant episode of self-vindication. And even if we are driven to ask, we cannot even pose the question as a hypothetical scenario, a *what-if* situation because, as mentioned above, one cannot think otherwise than what one is thinking of doing. Consequently, if asked why we should not just opt out of agency, we can never really make the decision to opt out, as the decision was never really ours to have (because this would be contradictory to the knowledge we have of being agents). Accordingly, Velleman falls flat into the trap of self-vindication.

#### 4.3.5 To Tally Up

Velleman’s constitutive account of agency is more successful than accounts offered by Korsgaard and Railton. First, with regards to norm-violation, in Velleman’s account, an agent is not at risk of diminished agency or termination of agency just because she violates a norm. Rather, the drive for sense-making, still intact, is left unfulfilled because those actions which violate norms may very well be those which fail to make sense to the agent. But, and more impressively so, it is possible for Velleman’s agent to violate norms in the strong sense by

reframing what is ‘unthinkable’ so that it becomes intelligible for the agent to do (as explained in § 4.3.1).

When it comes to the unavoidability of agency however, Velleman does not fare well. First, he makes it clear that the parts that are essential to agency, such as autonomy and understanding cannot be dissociated from yourself. But what makes his account a bit more palatable is that he explains why acting for reasons is essential to agency, instead of simply stipulating that this is just something we do. (See §4.3.2 and §4.3.3).

When asked the why-care question, Velleman responds that by asking this question to begin with, we are appealing to our drive for sense-making, and because of this, agency is self-vindicating. But what makes his response a bit more interesting than the other accounts is that if we have the conception that we are agents, then we cannot entertain a contradictory conception of shmagency.

Overall, I think Velleman does a better job at the Shtest than both Korsgaard and Railton and I attribute this success to the idea of seeing a drive as that which is constitutive of agency. The drive allows for non-arbitrariness in the following way: the drive is different from other desires and is something which is universally shared between agents.

#### **4.4 The Diagnosis**

In the previous and present chapters, I have examined three constitutive accounts of agency to determine if they are able to address the four different problems set out in the Shtest. I concluded that every account fails, some more severely than others. Here, I provide the diagnosis problem by problem as outlined in the Shtest, and briefly showing the failure of each

account. I then proceed to identify and discuss a critical flaw each account has committed to which might explain the failure of each.

#### *4.4.1 Feature 1: Norm-violation*

Korsgaard is unable to show how norm-violation is possible. If Korsgaard's agent deliberately violates norms, she either ceases to be an agent or her agency is severely diminished. This is obviously an undesirable outcome (See §3.2.1).

Similarly, Railton's account allows for weak norm-violation, but not necessarily strong norm-violation and consequently does better than Korsgaard. By showing how, when a person forms certain beliefs, that the belief somehow misses its mark, is to show weak norm-violation.

The account which performs best with regards to norm-violation is Velleman's. Under Velleman's account, weak norm-violation is possible because an agent may do things which are not intelligible to the agent, but this just means the drive for sense-making is left unsatisfied.

#### *4.4.2 Feature 2: Unavoidability*

Korsgaard's account commits to the notion of unavoidability in a strong sense by claiming that we are condemned to be agents and that we can never *not* act, for even an episode of inaction, is a form of action. Thus, under Korsgaard's account, agency is completely unavoidable (See §3.2.1).

This same failure is shared by Railton. Railton too buys wholeheartedly into the concept of agency as being unavoidable. He commits to the notion of agency as being a deep, undeniable fact of what we are and that to imagine non-agency, or even shmagency is a strange thing to be doing.

And finally, Velleman too shares the same failure. Velleman openly refers to the inescapability of agency and that we are condemned to write our futures. Furthermore, that the aim of self-understanding is inescapable, it is something you cannot help having. He falls hard against this problem.

#### *4.4.3 Feature 3: Self-vindication*

Korsgaard falls flat against the self-vindication problem when she claims that we cannot *not* adhere to rationality norms when acting. Asking why it is we should care to be agents, we are asking something which is an action of which we cannot but adhere to the very norms it is questioning (See §3.2.2).

Railton, unfortunately falls against the problem of self-vindication just as hard. For Railton, by asking for reasons to be an agent, you do so from the place of agency and second, you cannot possibly ask for reasons outside of agency.

Velleman also falls hard against the problem of self-vindication. By asking why it is we should care to adhere to norms, the agent responds that to ask that question is to do so in virtue of trying to make sense of what one is doing. Because of this, asking the question is self-vindicating.

#### *4.4.4 Feature 4: Non-arbitrariness*

Korsgaard is in the least likely position to show how the constitutive feature of agency (being self-constitution through action) is somehow non-arbitrary from other features and how it is that this constitutive feature guarantees a privileged normative-status. It just seems unlikely

that something like self-constitution (through action) is able to ground rationality norms (See §3.2.3).

Railton suffers the same fate unfortunately, but to a far lesser degree. He manages to show how beliefs which are constitutive of agency are at least less arbitrary than Korsgaard's self-constituting actions. He shows that desires, which yes, can be perceived as quite arbitrary, are impacted or informed by our beliefs – thus rendering beliefs non-arbitrary. The problem for Railton however is that the non-arbitrariness of this constitutive feature hinges entirely on the commitment he has made to the unavoidability claim and the claim of self-vindication.

Velleman recognises a drive which is shared by all agents. This gives him the upper hand regarding the non-arbitrariness of his chosen constitutive feature of agency. Because this is a drive the hallmark of agency, non-arbitrariness is established: Velleman is more specific and details that this thing which is constitutive of agency is a drive for sense-making.

#### *4.4.5 The Critical Flaw: Unavoidable Agency*

Where some accounts fail to pass the Shtest completely, some seem to get a small portion right. There's a certain trend shared by all accounts and it is this: every account fails, almost entirely, to deal with the problems of the unavoidability of agency and the problem of self-vindication. And there could be a very good reason for this. Perhaps, all of the constitutive accounts discussed so far link, too intimately, the notion of unavoidability and self-vindication. It could very well be the case that as far as you derive self-vindication from unavoidability, you will always find yourself failing the Shtest. Why?

Because you are always linking them in such a way that they don't allow you to then explain what is non-arbitrary and unique about your chosen constitutive feature. If you are going to



answer the why-care question by appeal to a constitutive feature of agency, you shouldn't make agency unavoidable in virtue of the very feature that answers the why care question. Constitutivists who so easily link up or collapse self-vindication and unavoidability fail the Shtest because they locate the unavoidability in the very feature that allows them to point fingers and say, you can ask that question because you are using the very thing you are doubting by asking it. Self-vindication should not come from unavoidability.

And so, the struggle to overcome the challenges set out in the Shtest is because constitutivists so closely link the notion of the unavoidability of agency and self-vindication. Overall however, the actual problem lies not just with the intimate link that has been made between these two, but really, the notion of unavoidability of agency.

When I talk about unavoidable agency, recall, I mean finding oneself an agent with no choice in the matter, not being able to do otherwise but be an agent, someone who is – to use Korsgaard's pithy expression – condemned to this role. Further, this agent is condemned to possess the things which are constitutive features of her agency. So, for instance, adhering to rationality norms may be a constitutive feature of agency and therefore, she has no other choice but to adhere to rationality norms. Agency is something this individual is thrown into, and because of this – is forced, by no choice of her own, to just do what agents do. It is because of this unavoidability of agency, I argue, that constitutivists are facing the other problems.

As I have already argued, the unavoidability of agency seems to be a necessary move for the constitutivist to answer the why-care question: we have no choice but to be rational given that this is what it means to be an agent; it is just something we find ourselves doing. Of course, this is a very strong strategy as has already been explained previously – but faces what appears to be an insurmountable set of challenges, which I grouped under the umbrella of the Shtest.

What happens if agency is an enterprise just like any other enterprise, such as playing chess? Because of this, one must have *reason* to be an agent as opposed to a shmagent, just as one should have a reason for wanting to play chess as opposed to playing shmess. What the unavoidability and inescapability of agency gives the constitutivist is an answer as to why we *just* are rational (why we are agents). But this also makes her incapable of answering the question of why we *should* be rational (why we should be agents). Hence, it seems that it is the constitutivist's unavoidability conception of agency that prevents her from explaining the normative grip of rationality norms.

However, if the unavoidable and inescapable nature of agency are the only source of this inability, then we can abandon them in favour of a voluntary conception. We then stand a better chance to pass the Shtest, and hence to ground rationality norms in agency.

Let me explain this through an example. Imagine you suddenly woke up one morning to find yourself climbing mount Everest. This is an involuntary occurrence and one you never planned for, nor have an idea now why it is that you find yourself climbing a mountain. You find yourself in a harness with an ice-axe in your left hand and rope in your right. You have no other option but to use the tools available to you to survive. You have no option, that is, but to be a mountain climber.

Within that, of course you have some choices. You can use the ice-axe to climb upward, or the rope to lower yourself downward, or to unclip yourself from your harness and let yourself fall. These are the options available to you. You may, at this moment, experience a semblance of free choice because you have three options before you. But what you don't have is a choice about being a mountain climber. As a mountain climber (something you have just found yourself being), you are expected to climb mountains, which means you should use the ice-

axe. Even if you decide to scale down the mountain, that's also something mountain climbers do – so either action will be that of a mountain climber.

The only escape is to let go of everything and fall to your death. You cannot legitimately ask this unavoidable and inescapable mountain climber any of the following questions: why be a mountain climber as opposed to a barista? Why use an ice-axe or harness or rope? Why *should* you use a harness or a rope or an ice-axe? The answer is simply going to be a cry of desperation: 'I cannot do otherwise!'

This is just how constitutivists conceive of agency. We find ourselves thrown in it; the only permanent escape is death; and even that final exit (if in the case of suicide) must be made by an agent. A dire consequence of this picture is that we can't then ground normativity in it: if agency is unavoidable, and adhering to norms is constitutive of agency, then choice has been extracted from the process of norm-adherence. But as I have argued, deliberate norm-adherence presupposes norm-violation. It follows then that the unavoidability of agency is the root of the problem of norm-violation.

Korsgaard commits to the notion of the unavoidability of agency when she claims that we are somehow condemned to be agents. Railton commits to the notion of the unavoidability of agency when he maintains that agency is just a deep fact about us and something which is inescapable. Velleman runs into the problem of the unavoidability of agency by claiming that the features constitutive of agency are features from which you cannot dissociate yourself from.

Even if some of the accounts can overcome the problem of norm-violation (and barely so), they still fall against the other problems, specifically the problem of unavoidability, self-vindication, and non-arbitrariness.

#### 4.5 Is This Just an Analytic-Philosophy Quirk?

It would not be amiss to wonder if the failure of the constitutive accounts I have examined so far is somehow linked to the fact that they are all from the analytic tradition in contemporary philosophy. To debunk this, it's worth to briefly discuss a constitutive account from the continental tradition and to show that this account experiences the same failures.

Katsafanas posits a constitutive account with a drive as the constitutive feature (similarly to Velleman). He identifies as the drive's aim the Nietzschean will to power. It's important to note that it does not matter to me if Katsafanas' account is plausible or not and I do not expect the reader's buy-in into his idea. I am also not necessarily concerned with problems that have been identified by others regarding his account – all I want to take from discussing this account is that the same trend is happening even under an account which falls under a continental framework. Thus, showing that the problems are not isolated only to those account that fall under an analytic framework.

What does the drive for power look like? Katsafanas argues that there is a constitutive aim of all action (again similarly to Velleman):

[I]f action has a constitutive aim, then that aim will be present in every instance of action. Thus, it will give us a non-optional standard of assessment for action, a standard that applies merely in virtue of the fact that something is an action. (Katsafanas, 2013, p. 630)

Katsafanas identifies this non-optional standard as the *will to power*. How exactly does the will to power manifest as this non-optional standard? Katsafanas explains that it takes the following form: all action, if aimed at power, would aim at encountering and overcoming resistance to ends. What does this mean?

The aim of the drive is, as he puts it, “the drive’s distinctive form of activity, that by means of which it is individuated. The object of a drive is a temporary goal upon which the drive finds expression” (Katsafanas, 2013, p. 636). Hence, an end can be seen as a kind of event in which the drive can be expressed. The aim however is activity-centred. If this is so, then drives are not so much goal-directed as they are process-directed.

The process Katsafanas identifies is to encounter and overcome resistance. For example, if you pick a crossword puzzle to complete, you will choose one which is not too easy for you to do. You will choose a crossword puzzle that offers the opportunity for you to overcome resistance (Katsafanas, 2013, p. 643). But, because the drive for power is process-directed, the satisfaction or success of the project is dependent on how challenging the project is. The attainment of the end is just an occasion for expression, not the fulfilment of this drive. For Katsafanas, this is how we are able to measure the success of an action: an action is successful insofar as it encounters and overcomes resistance.<sup>58</sup>

Katsafanas ties the idea of the drive for power to the notion of value. He does so in a rather interesting way. First, he establishes that power is the ultimate value – so much so that the drive to power cannot be reassessed, we are “inescapably committed to valuing power” (Katsafanas, 2013, p. 648). Furthermore, this drive is a value against which all other values must be measured or assessed. According to Katsafanas

facts about which actions we have reason to perform are determined not only by the degree of resistance that the potential actions afford, but also by facts about the relationship between potential actions and other values; however, these values must in some way be vetted by the standard of will to power. (Katsafanas, 2013, p. 646)

---

<sup>58</sup> Katsafanas makes several empirical claims of this kind that might be easily questioned, or at least we could ask for empirical reasons to believe them. But remember, my aim is not to endorse or criticize his view. Just to show that it equally illustrates my diagnosis.

But what happens in moments of conflict between the ultimate value (that of the will to power) and other values? It appears that the only way to get around this problem is to modify the other value so that it can align with the will to power.

The will to power as that which encounters and overcomes resistance. Identifying this as the source of normativity is, for me at least, largely problematic. There needs to be a set of principles which govern this type of resistance surely. What principles does this notion of resistance provide for us to assess other values for instance? So, while Katsafanas's position can provide a model for ranking reasons, in a way that Velleman could not, unfortunately the ranking model Katsafanas provides doesn't seem to be in the position to provide accurate or correct verdicts.

The idea behind Katsafanas's constitutive account of agency is to identify that which gives us most reason to perform those actions which best fulfil the constitutive aim. Because we strive for the best reason, it means that some reasons fall short of being the best way to fulfil the constitutive aim; yet sometimes we base our actions on those very not-so-great reasons. This allows for norm-violation in obvious ways: "the constitutivist about action will not be bothered by the fact that agents sometimes have reason not to perform the actions that best fulfil the constitutive aim" (Katsafanas, 2013, p. 653).

Upon a very superficial rendition of his account, Katsafanas allows for norm-violation. A feat accomplished by both Katsafanas and Velleman. It should not go unnoticed at this point, that the only two accounts which can meet the norm-violation challenge are those which posit drives as being the constitutive feature of agency (this is something I discuss in Chapter 6).

But, just as the others do, Katsafanas blatantly commits to the unavoidability of agency by claiming that the drive for power is inescapable: "Nietzsche is grounding power's privileged

evaluative status in an *incapacity*: it is the one value we cannot give up, insofar as we are engaged in willing” (Katsafanas, 2013, p. 634). The consequences are as follows:

By identifying a drive to power which is unique to agency, Katsafanas attempts to distinguish this drive from other contingent desires. For him, this drive is different from other desires in that it is what motivates a person to perform any action at all, as opposed to other motivations which explain the choice of certain actions over others. Because this drive is constitutive, it is unique. But this is similar to the argument Velleman tried to run – but, this is problematic because it depends on the unavoidability argument to work.

Unfortunately, though, when it comes to the problem of self-vindication, it should be obvious that this account is no different from any of the others. In answering the why-care question, Katsafanas might say we should care to be agents because we have reason to be an agent as opposed to a shmagent in virtue of us having this drive, which automatically runs up against the problem of self-vindication.

#### **4.6 Conclusion**

At this point, it should be clear that the four constitutive accounts I have discussed so far do not do well with regards to the Shtest. In the previous section (§4.4), I provided a diagnosis and I went over the accounts again and highlighted their successes and failures. Because of this, I will not provide further summaries in this conclusion but will keep it short. Suffice to say that the key failure experienced by all accounts, I have argued, can be pinpointed in the intimate relationship that has been established between the notion of the unavoidability of agency and that of self-vindication. In the next three chapters, I focus on my positive contribution.

# Chapter 5 – Elective Agency

## 5.1 Introduction

In the previous two chapters, I discussed four constitutive accounts of agency, which met with varying degrees of success with regards to the Shtest. The first two accounts I focused on were those offered by Korsgaard and Railton. Korsgaard argued that adhering to rationality norms is constitutive of action, and further, that action is constitutive of agency. Railton argued that adhering to rationality norms is constitutive of belief, and further, that belief is constitutive of agency. Both accounts try to ground the source of rationality norms in a constitutive feature of agency. I argued, however, that in order to successfully ground the source of rationality norms, a constitutive account must meet the four challenges set out in the Shtest. I argued that neither Korsgaard, nor Railton could successfully meet these challenges.

The remaining constitutive accounts I focused on were put forth by Velleman and Katsafanas. Velleman tries to ground the source of rationality norms through a different approach, by foregrounding the role of a special drive as constitutive of agency. Even though Velleman's account did a far better job at addressing the problems set out in the Shtest, it unfortunately still fails to meet all the challenges. Katsafanas too, grounding his theory in the continental tradition, posits a drive in which to ground the source of rationality norms. He too fails the Shtest.

In this chapter and the next, I offer my unique contribution to the constitutivist debate. I propose a constitutive account of what I call elective agency, and show how even a black-box version of it fares better with the Shtest in virtue of introducing *choice* about agency. The novel claim I defend is that we don't just fall into agency and are then forced to play the game of agency. Rather, we choose to be agents and constantly choose to play the game of agency. I further



show that it is because of this *elective choice*, that an account of elective agency can sustain the normativity of rationality.

This chapter is comprised of five sections. In section 5.2, I set the stage for my account. In section 5.3, I discuss a theory called Hierarchical Voluntarism put forward by Ruth Chang (Chang, 2009). I use ideas from Chang's theory as a foundation for my constitutive account of elective agency. In 5.4, I introduce a black-box notion of elective agency in terms of its functional role. In 5.5, I conclude.

## 5.2 Setting the Stage

According to Enoch, the source of rationality norms cannot be found in what is constitutive of agency. The reason is that there is very little chance of showing successfully why agents should care about norms at all unless we can also show why they should care about being agents.

All of the constitutive accounts I have looked at so far have wholeheartedly committed to one foundational idea: that agency is unavoidable.<sup>59</sup> This is important to constitutivists, because it allows them to make room for grounding normativity in agency. For in the process of caring about norms, one does so from a place of agency, which is unavoidable. So then, if you are an agent, you care about these norms by default.

The problem here is twofold. First, moving from the premise that agency is unavoidable, we have set up too intimate a link between the unavoidability of agency and the problem of self-vindication. Insofar as you derive self-vindication from the unavoidability of agency – you will

---

<sup>59</sup> When I use the term unavoidable agent or unavoidable mountain-climber, I adopt it as a term of art to apply to agents – X is something we just are and cannot opt in or out of.

never be in the position to pass the Shtest.<sup>60</sup> But second, there just seems something intuitively wrong about arguing for the unavoidability of agency in virtue of asking the why-care question – or to be more specific, there is something wrong with arguing for the unavoidability of agency whilst in the process of exercising agency.

What happens if we knock out that foundational commitment upon which the constitutivists depend. What happens if we play around with the idea that agency actually isn't all that unavoidable – that it's somehow *voluntary*? What does this mean for the debate?

It means that Enoch can no longer claim that grounding normativity in constitutive accounts of agency is impossible. It also puts us in a far better position than any of the other constitutive accounts to pass the Shtest. In section 5.4, I introduce an account that does this. But first, I introduce a theory proposed by Chang (2009) which I use as the foundation for my account.

### **5.3 Hierarchal Voluntarism**

Normative voluntarism is the position that normativity can be grounded in an agent's will. According to Chang (2009, p. 244), normative voluntarists locate the source of normativity in us, but not in our passive states. Rather, normativity has its source in something we do, and, in particular, in our active attitudes of willing or reflective endorsement. By willing something, that is, by actively engaging our volition, we can give a consideration the normativity of a reason.

The way I understand voluntarism, particularly Korsgaard's (1996) for instance, is that one constitutes oneself through actions which one essentially wilfully endorses – in this way

---

<sup>60</sup> At this juncture, I'd like to re-emphasize that as far as self-vindication is derived from the unavoidability of agency, it is problematic. It may very well be the case that self-vindication, on its own (not tied to the unavoidability of agency), might not be problematic. This is something I argue for later on in this chapter.

grounding normativity through the agent's will. However, Korsgaard's account is also very much constitutivist (as I have already discussed in chapter 3). It is possible to merge voluntarist positions with constitutive positions, most notably this kind of marriage between the two can be found in Korsgaard (1996) and Bratman (2007). Bratman maintains that

the ground of agential authority involves higher-order attitudes whose function includes the constitution and support of the temporal extension of agency [...] the ground of agential authority involves higher-order intentions, plans, and policies. (Bratman, 2001, p. 320)

This higher-order intentions and plans involve, very closely, the agent's will: without this will, these intentions and plans would not come into existence.

There are many concerns with the voluntarist position, and as has been noted by Chang, it is considered the "runt of the litter" (Chang, 2009, p. 245) in comparison to normative internalists and normative externalists (see §2.4.1). Why is this so? Because it seems rather strange that an agent is somehow able to imbue a choice, decision or consideration with normativity simply by exercising an act of will. Of course, many theorists (Hobbes, 1651; Kant, 1785; Korsgaard, 1996) have taken up the challenge and have put forth innovative arguments to show how this can be the case.<sup>61</sup> However, the most notable challenge facing these voluntarists is that "if the source of normativity of our reasons is in our willing something, then practical reasons become objectionably arbitrary" (Chang, 2009, p. 247). This means that the voluntarist must show how imbuing a choice with normativity is not objectionably arbitrary. This is something Chang attempt to do in Chang (2009).

She proposes a 'hierarchical voluntarism' which denies the premise that "normativity has a univocal source" (Chang, 2009, p. 246). According to Chang, all voluntarists hold that some

---

<sup>61</sup> I won't be going into those theories here for lack of space.

reasons are grounded in a normative source of the will. She takes note of the objection mentioned above and introduces a kind of hybrid voluntarism where normativity is dependent on an act of will, while claiming that there are some instances where not all reasons are voluntarist. She claims that these voluntarist reasons play a role *only* when non-voluntarist reasons have been depleted. Non-voluntarist reasons, according to Chang (2009, p. 246), are those “reasons whose normative source is not an act of will” – examples of this include involuntary drives, desires and other such agential features (essentially what constitutive accounts are pushing for).

Her argument looks like this:

P1<sup>Chang</sup>. Through an act of will, we generate reasons.

P2<sup>Chang</sup>. To generate a reason is to endow that reason with normativity.

C<sup>Chang</sup>. Through an act of will, we generate normativity.

I discuss each premise.

### 5.3.1 *What Can We do Through an Act of Will? (P1<sup>Chang</sup>)*

Chang asks us to imagine that we are faced with two possible career choices. First, we could choose to become a philosopher and spend our days writing papers, reading articles and contemplating the meaning of life; or, second, we could become a trapeze artist and spend our days wearing leotards, chalking up our hands and performing daring leaps under the big top (Chang, 2009, pp. 249 - 250).

She asks us to further imagine having weighed up every considerable reason for choosing either. Suppose that after considering every possible reason to choose either the career of a

philosopher or that of a trapeze artist, we find ourselves unable to reach a decision, because the reasons weigh equally on either side. Perhaps, “you have sufficient reason to choose among several alternatives or [your] choice is beyond the reach of practical reason” (Chang, 2009, p. 249). She rightly points out that this kind of decision is a big one as it deals with your future career – something that will presumably last a lifetime and inform the kind of person you will be in future. “What should you do? It seems it would be a mistake for you to simply pick or plump for one career” over the other (Chang, 2009, p. 250).

When our reasons have ‘*run out*’ in this manner, we are left with three possible options. First, we do what Chang has pointed out would be a mistake to do, we simply pick one career – perhaps we refer to a coin toss to do so. Note, Chang is not saying that there are never instances where we randomly pick when our reasons have run out. For instance, if our reasons have run out between two possible meals for our dinner, then picking seems appropriate. However, when it comes to the big decisions, it is best not to just randomly pick. Second, we go back, and we continue to examine and deliberate over the reasons, until hopefully *something* pops out at us as a reason that might tip the scales to choose one career over the other. The reason, Chang identifies, we continue to deliberate over the reasons, even though we know they have run out, is because we don’t *really know* they have run out, not in the sense for us to stop with the search (Chang, 2009).

A way around this, for Chang (2009, pp. 250 - 251), is that we may not be epistemically certain that our reasons have run out but can be *practically* certain. If you have this kind of practical certainty, then to continue to deliberate over the reasons is actually a rather irrational thing to do. So then, if our reasons have run out, we can either pick randomly or we can continue with our search. Both instances are not what we should be doing. There is a third option she proposes.

Chang (2009, p. 253) identifies two stages in the deliberation process. The first stage is where we consider all options and we reach the conclusion that we are practically certain that there is no point for further deliberation because our comparative and non-comparative reasons have run out. There could be, according to Chang, a second stage where we can deliberate further. At this second stage, we could *will* an alternative.

What would this look like?

After examining every possible reason to either become a philosopher or a trapeze artist, I am left dumb-founded. What I do not do is plump for a reason or continue deliberating. Instead, I *will* a consideration to join the circus to be a reason, and by doing this, I

give the reasons in favour of the circus career extra normative force. The extra force of these reasons can then give you most reason to choose that career. (Chang, 2009, p. 255)

But, how can willing something give us more reason to do it?

Chang distinguishes between voluntarist and non-voluntarist reasons. Non-voluntarist reasons (or given reasons<sup>62</sup>) are those which “we ordinarily take ourselves to have – reasons whose normativity derives either from normative reality or from our desires, but not from our own act of will” (Chang, 2009, p. 256). Conversely,

Our voluntarist reasons [...] are the reasons we create for ourselves by taking a consideration to be reason when our given reasons have run out. Thus when your given reasons for choosing one of the careers run out, you can take a consideration in favor of one of the careers as normative for you, thereby creating a new, voluntarist reason to choose that career. (Chang, 2009, p. 256)

---

<sup>62</sup> It should be noted that some theorists take issue with the notion of given reasons or givenness. I will not engage in this debate as it would take me away from what I am hoping to accomplish here. But, for more on given reasons and givenness, please see MacDowell (1994); Schiller (2007); Sellars (1956; 1962).

Chang asks us, for simplicity sake, to imagine that there are only two given reasons to consider with regards to either becoming a philosopher or a trapeze artist. After examining the reasons, you are practically certain that you have run out of reasons. Chang proposes that you could take the secret love of wearing sequins as a reason to choose to become a trapeze artist rather than a philosopher. “Through an act of will, [you can make it] a reason that [is] relevant to your choice” (Chang, 2009, p. 257).

Sometimes, and importantly so, contents of given reasons can also be the contents for voluntarist reasons. What this means is that sometimes, when we have effectively run out of reasons and we are practically certain of this, we go over the reasons again. If we find ourselves going over the reasons again, even though we are practically certain we have run out of reasons, it isn't irrational to agonise over the given reasons if we intend to deliberate “over whether and which such considerations to create as new, voluntarist reasons for yourself” (Chang, 2009, p. 258).

Creating a voluntarist reason requires a certain buy-in from the agent that this consideration – not any other – is “a consideration that counts in favour of some action or attitude” (Scanlon, 2006, p. 231) – in particular, *this* reason I have *willed* as *my* reason counts in favour of a particular action. This investment I make means that this reason is of some value to *me*. We can think of the distinction between given reasons and voluntarist reasons in terms of objective and subjective value. There are reasons which are objectively valuable insofar as they are readily applicable to all in a given situation. However, there are voluntarist reasons which are subjectively valuable insofar as they are valuable to a particular individual in a particular situation. Whatever Chang herself means by the given-voluntarist distinction, I think this is the most plausible reading of it.

Note that voluntarist reasons are different to plumping for a reason where there is little investment from an agent. If someone asks me to choose between a black or blue pen to sign a document, I particularly don't care which pen I use as long as both are permissible to use. This would be an instance where it doesn't matter which I end up with. I'm not invested in either. Here, I randomly pick. Perhaps I am signing an employment contract and, in this instance, it's still permissible if I use either colour pen. Maybe I have equal weighting of given reasons to choose either, it doesn't matter. But then suppose I remember my father saying that signing in black looks more official. So, given that this is an important document for me to sign, a very official one at that, I will take the consideration of my father sagely nodding his head and telling me about signing in black to be the reason that I want to sign this employment contract with a black pen. If it turns out that I mistakenly sign with blue ink, this will upset me.

If I can generate reasons through an act of will, does this mean I will automatically act on this reason? In the next sub-section I look at what happens when we generate reasons.

### 5.3.2 *What Happens When We Generate Reasons?* (P2<sup>Chang</sup>)

By taking a consideration as a voluntarist reason, we add normative weight to this reason. We now have more reason to do the thing we have voluntarist reason to do – whereas before, we had run out of our given reasons. “This willing creates normativity by creating new reasons whose normativity derives from the very act of will” (Chang, 2009, p. 255).

How does this happen? According to Chang, it's not the alternative that is somehow imbued with normativity, but rather that the act of willing is what adds weight to the reason to favour



one alternative over another.<sup>63</sup> The fact that we have the capacity to will and that we use it in this way is why the reason is imbued with additional normativity – something that a given reason does not have when reasons have run out. This is the important distinction between willing a consideration to be a reason over plumping for an alternative. Through the act of willing, I am committing to the alternative as something I care about (this is something I explore in 5.4.3) – it takes on normative force for me.

But how is it that we have this capacity and what does it mean?

### 5.3.3 *Why and How do We do This?* (C<sup>Chang</sup>)

T. M. Scanlon gives a nice formulation of the worry I have in mind (in the context of Raz's work):

if the normative-status of counting in favour of acting a certain way is something that certain considerations can just have, how can it also be something that we can confer on certain considerations by our choice of ends? (Scanlon, 2006, p. 231)

A way to answer this is to appeal to what Chang calls a “rational identity” (Chang, 2009, p. 259). She argues that

unlike the sheer willing of an alternative, willing a consideration to be a reason is part of the process of making oneself into a distinctive normative agent, that is, creating one's own ‘rational identity’. (Chang, 2009, p. 259)

A rational identity, according to Chang, is what each individual agent imagines their perfect, ideal rational selves would look like in this world. Importantly, your rational identity may be quite different from mine. For instance, it makes most sense for me to choose to be a trapeze artist over a philosopher as this is what I have most reason to do – especially when I am the

---

<sup>63</sup> In the previous chapter (§4.3.3), I noted, as a concern for any constitutivist, that there must be a way to rank reasons. Answering the question of how we choose which considerations to take as our voluntarist reason, may be one way of explaining how reasons are ranked.

one who can create the very reasons in forming my rational identity. Perhaps you think it's best to generate a reason to be a philosopher instead and that is what works best for your ideal rational self.

As Chang claims,

if creating reasons for oneself is making oneself into a distinctive normative agent, then the difference between the dinners and career cases can be explained by the role such choices play in one's rational identity. (Chang, 2009, p. 259)

What isn't clear, though, is how we are both responsive to reasons (in the first part of the deliberation process when we're deliberating over our given reasons) and have the capacity to create reasons (Chang, 2009). To answer this, Chang defines rational identity in terms of a perfect rational self. "It is a function of the reasons that determine what you have most reason to do in actual and hypothetical choice situations" (Chang, 2009, p. 260). In taking certain considerations to be voluntarist reasons, once my given reasons have run out, I create reasons which support my rational identity.

Perhaps an anti-voluntarist (e.g., Cohen, 1996) could say: 'hold on a moment, these voluntarist reasons sound quite arbitrary. It seems that an agent can will any random consideration to serve as a reason. Because of this, we cannot say these reasons are normative'. The response put forward by Chang is that these voluntarist reasons which hold a privileged normative-status are distinctly non-arbitrary, because "the hierarchal relation between one's given and voluntarist reasons guarantee that we can never create a voluntarist reason that goes against our all-things-considered given reasons" (Chang, 2009, p. 270). Thus, our voluntarist reasons are restricted and guided, one could say, by the given reasons – ensuring normativity.

If, after examining my reasons for becoming a philosopher versus my reasons for becoming a trapeze artist, I realise that I have run out of reasons to choose one or the other. I might think

back to a philosophical problem or thought experiment I found most intriguing, maybe Frank Jackson's Knowledge Argument (Jackson, 1982). I might take this consideration as normative reason for myself and as a consequence, I take it as a voluntarist reason to choose a career in philosophy. By doing this, according to Chang (2009, p. 263), I "exercise genuine agency in 'making'[myself] into [an] agent with one rational identity rather than another".

#### 5.3.4 *The Take Away*

According to Chang, when we have run out of reasons to choose one thing over another, we are able to will a consideration to be a reason. It can be in one of two ways: i) we will a new consideration (or a non-comparative reason) to be a voluntarist reason to make a decision, or ii) we will a given reason to be a voluntarist reason to make a decision. Using ideas from Raz's "Incommensurability and Agency" chapter (Ch 3) I explain further how we are willing a consideration to be a voluntarist reason (Raz, 1999).

Meet Julian. He's a mild-mannered accountant and works at a bank. During his lunch hour, Julian sneaks to his car which is parked in the lowest level basement. There, he sits staring at the wall for the full hour. Julian tells himself that he is meditating. But he isn't. He is trying to figure out if he should divorce his wife and commit his life to becoming a world-renowned Ping Pong player or stay with his wife and become a better accountant.

He thinks of the reasons to become either and he realises that he has run out of given reasons. The given reasons for divorcing his wife and becoming a Ping Pong player are that he really dislikes his wife; he really loves the sound a Ping Pong ball makes when it hits the table (among many other reasons). He remembers being really good at the game at university. The given reasons for staying with his wife and becoming a better accountant are: His wife is a good

person and he doesn't want to hurt her; playing around with numbers as skilfully as he does makes him feel smart; He remembers being really good at numbers at university.

Now let me borrow a distinction from Joseph Raz, in order to make my point. Raz contrasts two conceptions of action - the rationalist and classical.<sup>64</sup> The rationalist theory holds that an agent performs the action which is best supported by the strongest reason (Raz, 1999, p. 47). Conversely, the classical conception sees an agent taking an action because the agent chooses to do so. The distinction here is important. Under the rationalist conception, we act in accordance with the best possible reason (thus, the reason informs our choice and has normative force). Under the classical conception, we take reasons to be those things that render "options eligible" (Raz, 1999, p. 47). The second important distinction is that the rationalist view sees an agent's desire as being a reason to act and the classical conception as regarding the will as "an independent factor" (Raz, 1999, p. 47). The third important distinction is that the rationalist conception does not account for incommensurability and sees this as a rare phenomenon, where the classical conception can account for incommensurability.

In the above example, Julian is unfortunately facing an incommensurability of reasons. He has run out of reasons to choose either to continue being an accountant and stay married to his wife or to become a world-class ping-pong player. So, he wills a new consideration to be a reason for him to choose one over the other. According to Raz,

the will is the ability to choose and perform intentional actions. [...] Commonly when we so choose, we do what we want, and we choose what we want, from among the most eligible options [...], similarly, when faced with unpalatable but unavoidable and incommensurate options, [...] it would be correct to say that I want to give up the one I choose to give up. (Raz, 1999, p. 48)

---

<sup>64</sup> I admit that this is an oversimplification of the discussions and debates out there concerning the distinction between classical and rationalist perspectives of human agency. Due to space, I cannot elaborate on it much. For more on the distinction between classical and rationalist perspectives, see Bratman, 2006; Raz, 1999; Stocker, 2006.

In this instance, Julian wills the reason that he is skilful with numbers to be the reason to choose to continue to be an accountant. He exits the car, goes upstairs and continues with his job.<sup>65</sup>

Now recall Chang's claim that by willing a consideration to be a voluntarist reason, we imbue that reason with normativity, such that that reason becomes a reason for us to do one thing over another. When Julian takes the consideration of the fact that numbers make him feel smart to be the voluntarist reason for him to do something, he is imbuing that reason with normativity, such that he now has a reason to continue on as an accountant, whereas before this, he did not have this reason to choose one over the other.

How is Julian's willing a voluntarist reason different from plumping for a reason in this example? Imagine Julian, whilst sitting in his car, is thinking about the many reasons he has to be an accountant versus those for becoming a famous ping-pong player, and he realises that he has run out of reasons to choose either. But he knows a decision must be made. He can't just sit in his car for the rest of his life. This becomes burdensome for him. So, instead, he says 'if the next car that drives past is red, then I'll be a ping-pong player'. The next car that drives past is blue – Julian continues to be an accountant. There is something arbitrary about this decision-process, especially since it's such a big decision. But what this means is that Julian has very little reason to *want* to be an accountant if the decision was reached through a random colour of the next passing car.

So, the difference between willing and plumping is this. When I will a consideration to be a reason for me to do something, I actively invest in that consideration as something of value to me. When I plump for a reason, then *any* consideration will do – I'm admitting that I don't

---

<sup>65</sup> For more on how willing is an act of choosing or is the ability to choose, see Bratman, (2007); Frankfurt (1971; 1988; 1999; 2004); Raz (2011)

necessarily care about the outcome either way. Accountant or ping-pong player... who cares? But one should care. This a big decision and one which will inform what kind of rational identity you have.

Are all choices the same? I argue that some choices are worth far more than others. For instance, the choice to purchase a Coke rather than a bottle of water intuitively seems to be worth less than choosing which partner to marry in life. Let's call the process of choosing the Coke a lower-level choice and the process of picking whom to marry – higher-level choice. Other examples of lower-level choosing could be whether or not to buy lunch at work or make lunch at home; which brand of toothpaste to pick; which grocery store to purchase apples from. Examples of higher-level choosing could be whether to choose Cambridge or Oxford for a post-graduate degree; whether to purchase a home cash or with a mortgage agreement; whether to abort a foetus or carry to full term. More often than not, it should be clear to the decision-maker whether or not the choice is a lower-level one or a higher-level one.

Maybe it could be argued that it's not always clear, because some choices may appear to be lower-level but have life-changing consequences. For example, choosing a brand of condom may be considered lower-level until picking a cheaper brand results in pregnancy. Or, picking the expired milk carton and deciding to drink the milk anyway, resulting in severe life-threatening illness. But the lower-level choosing and higher-level choosing isn't always necessarily about the possible outcomes, but rather about the value of the choice itself.

Maybe a utilitarianist might say, 'hold on a moment, but surely the choices are not as important as the consequences they bring about'. I'm not denying that consequences are important and do probably feed into our decision-making process quite a bit. I am, however, saying that there is something special about having the capacity for choice itself. Especially choice that is related

to higher-level decision making rather than lower-level decision making. Choosing between still or sparkling water when I don't have a particular preference is not an important choice for me. This is a lower level choice and, if we consider Chang, this particular decision-making process (if we ran out of reasons to choose either) can be satisfied with randomly picking one over the other. In lower level choosing, we value the things we are choosing just a little less.

So, for example, say I must choose between two different meal options on a flight: chicken or fish. I actually don't mind either. It makes no difference to me which it is I end up having. I say to the flight attendant: 'surprise me'. When I do this, I don't agonise over the decision nor will I feel delighted or disappointed in the outcome of this choice – and, if I did, it means I secretly wanted one over the other after all. This kind of flippant arbitrary choosing is not where one would ground the source of rationality norms.

But there is also higher-level choice – in some instances, even choice which could be considered to hold an ultimate value. What would this look like? For one, being faced with the big choices in life, such as which partner to marry or whether a person should opt for euthanasia rather than struggling through a difficult illness. It must be noted, however, that sometimes lower-level choices can be symptomatic of higher-order choices – think of choosing a salad rather than a roast beef pie because you're an animal rights activist.

As we increase the level of choice (based primarily on importance), it is possible that we increase the chances of grounding something as important as normativity. And again, going back to Chang, some *given* reasons, or non-voluntarist reasons hold normativity and that's why we opt into choosing accordingly (following the rationalist conception – see §5.4.2). But, considering the classical conception discussed in Raz (see §5.3.3), and the claims made by Chang concerning voluntarist reasons, it is possible that we create voluntarist reasons that

“render options eligible” (Raz, 1999). We don’t, according to Chang, create or generate voluntarist reasons if the options available to us are completely arbitrary – such as which meal to choose – this is an instance of lower level choosing and so this is probably not where we should be looking to ground the source of rationality norms.

But, regarding the bigger decisions in life, such as which career to choose, we have either given or non-voluntarist reasons to do so or we generate voluntarist reasons to choose one over the other, supposing we run out of given reasons to do either. There comes a point where we can indicate that a high-enough level of choice will be where one can ground normativity.

Often, such high-level choice happens, I argue (building on Chang), when reasons have run out and when we *will* voluntarist reasons<sup>66</sup>. Chang has argued that there is no reason to create voluntarist reasons when we have run out of reasons to choose between two meals. It’s best to just pick one. This is what I consider a lower-level choice. But when it comes to situations where our given reasons are not sufficient and we have no idea what it is we should be choosing – say, choosing between a trapeze artist and a philosopher as a possible career choice – this is a far bigger decision than choosing between two meals, the choosing that happens here is high-level choice.

And again, if we buy into Chang’s idea that willing a consideration to be a voluntarist reason is to generate or create normativity and if we further buy into Raz’s idea that willing is the capacity to choose, then high-level choice itself can be seen as generating normativity. Let me explain. When a person is deliberating about which career to choose, either to become a trapeze artist or a philosopher, it is possible that they run out of given ideas. A person then wills a

---

<sup>66</sup> Of course, I am not denying that high-level choice also occurs with given or non-voluntarist reasons. What I am arguing is that voluntarist reasons are yet another good place to ground the source of rationality norms especially because arbitrarily picking or plumping for one option over another does not happen with regards to voluntarist reasons.



consideration to serve as a voluntarist reason. This is an instance of high-level choice and an instance of imbuing a reason with normativity. It is now the case that because we have willed voluntarist reasons to choose one thing over another, we should be a trapeze artist rather than a philosopher.

Of course, choosing in accordance with given reasons or non-voluntarist reasons can also be high-level choosing. One might have plenty of comparative reasons to choose to be a trapeze artist over a philosopher and will make the choice accordingly. The choice itself is a higher-level choice and, arguably, because we are willing (making a choice), we are endowing this reason with normativity.

But, and this is important, given reasons are always either about low-level or high-level choices; however, voluntarist reasons are almost always high-level choices. Remember, that if it comes to running out of reasons, if there are two arbitrary options before you, it's perfectly acceptable to randomly pick. Actually, in order to abstain from further deliberation which can be considered a waste of time, one *should* just randomly pick between two arbitrary options. However, when we run out of reasons with regards to really big decisions, this is when we *will* voluntarist reasons. It's almost always the case then that voluntarist reasons involve higher-level choices. With this said, I now introduce Elective Agency.

#### **5.4 Setting the Foundation for Elective Agency**

In this section, I propose a black-box<sup>67</sup> notion of elective agency. By black-boxing, I provide an account in its overarching, generic form without showing *the mechanisms* behind *how*

---

<sup>67</sup> “The word black-box is used by cyberneticians whenever a piece of machinery or a set of commands is too complex. In its place they draw a little box about which they need to know nothing but its input and output.” (Latour, 2000, p. 681).

elective agency happens (as this is the focus in the next chapter). Instead, I argue that this black-box notion of elective agency could have good chance of overcoming the four challenges of the Shtest – thus effectively grounding the normative grip of rationality norms.

In this chapter, I am after showing two things. First, we are things which choose to be agents; second, what happens when we choose agency over non-agency or shmagency. I do this in two steps.

First, I show a constitutive link between choice and value-conferral.

Second, building on the previous two steps, I establish a constitutive link between high-level choice and agency. The steps look like this:

Step 1: High-Level choice is value-conferring. When we make high-level choices, we are conferring value through that choice.

Step 2: High-level choice is constitutive of elective agency. When we choose something as important as elective agency, we are conferring value onto it, and thereby imbuing it with normativity.

This, I argue, answers the why-care question.

### **5.5 Step 1: High-level Choice is Value-conferring**

I begin this sub-section with a very brief discussion on value.<sup>68</sup> There are various views one might adopt, such as realism, dispositionalism, and Fitting Attitudes (FA). For the realist, value

---

<sup>68</sup> I would prefer not to engage (at least not too heavily) on theories about value as this is too vast a literature for me to do it justice here. For what I have not covered here, please see Anderson (1993); Brentano (1969); Dancy (1993); Hansson (1998); Korsgaard (1996).

exists outside of the way humans perceive it. So, for instance, a red ball has the shape it does in virtue of it being spherical and the colour red is red insofar as light refracts from it. Dispositionalism is different in that the colour red is that which “has the power to arouse [red] appearances to normal human observers in standard conditions” (Jacobson, 2011, p. 1).

Fitting Attitudes analyses, on the other hand, “propose to analyse value, or some limited range of values, in terms of evaluative attitudes endorsed as fitting – or, alternatively, as appropriate, correct, merited, proper, rational, or warranted” (Jacobson, 2011, p. 1). For the way I use value here, I prefer to adopt the FA view, in particular the “*response-dependent account of value*” which sees value as “partly determined by human responses and attitudes”<sup>69</sup> (Jacobson, 2011, p. 5).

Why this view in particular? Because, as Raz puts it,

aspects of the world are valuable. That constitutes reasons for action. Because we are rational animals, ones with the power of reason, we are able to conduct ourselves in the light of those reasons. Being rational is being capable of acting intentionally, that is, for reasons, as one takes them to be, and that means in light of one's appreciation of one's situation in the world. (Raz, 1999, p.1)

Intuitively, because we are humans engaging with others within the world, it just seems likely that we value that which aligns with “one's appreciation of one's situation in the world” (Raz, 1999, p. 1). This works best with the Fitting Attitudes Analyses.

What I push for in this sub-section is that we have a unique ability to confer value when we choose. Imagine, while on my flight, I am offered one of two meals: chicken or fish. Maybe I'm not too picky and I don't have a strong preference for one over the other. Regardless, I *choose* chicken instead of fish because I like the smell of chicken more at this time. If I open the container and realise that instead of chicken, the flight attendant has handed me fish, I might

---

<sup>69</sup> For more on response-dependent accounts of value, see Wiggins (1987).

be slightly annoyed, a fitting attitude to have. Maybe for a couple of different reasons: i) the fish is smelly and I prefer the smell of chicken; ii) by choosing the chicken, it means I want the chicken more than the fish. It matters to me now, in virtue of my choosing it, if I get what it is I chose. Of course, this is low-level choosing with low value.

There may be some objective value in choosing fish over chicken and denying or affirming that objects have objective value is not something I wish to do here. But what I'm interested in is the value it has *for me*. I don't care if the person sitting two rows behind me chooses the fish or the chicken. His choice makes no difference to me. His choosing fish over chicken means that fish is what he wants. Fish is what he chooses.

There is an expectation that if you choose something, you will experience an outcome of that choice. Furthermore, that if you choose something, you confer value through that choice. And this is not to say that when I choose something, I confer value onto it that it did not have before, but it is to say that the object becomes valuable to me insofar as it is something I have chosen. And again, this can be measured against the lower-level and higher-level choice criteria. Let me explain.

In lower level choosing, I might choose a coke rather than an ice-cream. For me, it doesn't really matter which coke it is that I pick up off of the shelf and it probably wouldn't matter to me much if my friend snatched the bottle from me and replaced it with another. So, the value conferral that happens at this level is a lot lower than if I am choosing whether to marry Michael, Mason or Mary. This high-level choice becomes immensely important and is most certainly very much value-conferring. Even if marrying either Michael, Mason, or Mary may appear to depend on features such as their characters, my sexuality etc., We still, through the choice to value that person, confer value on the decision and the outcome.

With regards to choice being value-conferring, we can go deeper still. Two things: if we are able to imbue reasons with normativity through choice, then choice is also value-conferring if we can agree that normativity is valuable (I hope I don't need to convince anyone of this, given that this entire project is about how it is we should care about normativity); second, because choice is value-conferring, having the *capacity* to make high-level choices is also valuable. What do I mean by this?

Let me explain it like this. If I do not choose, then I do not confer value. This is not hugely problematic for an agent because there is some value that objectively exists without value-conferral taking place, such as the source of given reasons. However, it is problematic insofar as an agent is that thing which performs intentional actions. To do so, under both the classical and rationalist conception (See Raz, 2009), we perform intentional actions in accordance with reasons. Under the rationalist conception, we perform intentional actions (and thus choose) in accordance with what we have most reason to do, and under the classical conception, we perform intentional actions in accordance with what we will to be a reason. Under either view, we are choosing. Without this choosing, we would not act. Without intentional actions, we are not agents. This is why choosing has value (low-level choosing has a low value and high-level choosing has a high value). So, choosing is valuable and choice is value-conferring.

## **5.6 Step 2: High-level Choice is Constitutive of Elective Agency**

Now, the main claim of my argument. We *choose* agency. I do this in the following way. First (§5.5.1), I briefly set out what the pre-choice situation looks like. In §5.5.2, I argue that just like we run out of reasons to choose between becoming a philosopher or a trapeze artist, and

just like we are able to generate voluntarist reasons to choose one over the other, we too can run out of reasons to choose between either agency or shmagency.

Second (§5.5.2), I argue that this ability to generate voluntarist reasons affords us the opportunity to choose agency over shmagency. Importantly, this is different from Chang's view in a significant way: Where Chang's account claims that we make ourselves into agents with particular rational identities, I am claiming that we *choose to be agents*. When Chang says that we create our rational identity, agency is inferred as being present already. Her view is not a constitutive one. But I imagine there is a level of dependency going on between agency and rational identity, where the rational identity is dependent on the already established agency.

Third (§5.5.3), I argue that accepting rationality norms is constitutive of choice and that choosing agency is constitutive of elective agency. If this is so, then something like elective agency allows for a better constitutive account than any other account explored in this project.

#### 5.6.1 *The Pre-Choice Situation*

When Enoch asks why we should care to be agents rather than shmagents, I imagine that he pictured, in a very literal sense, a situation where a choice between the two is conceptually possible. Of course, with the incoherence problem, as well as the circularity charge pointed out in chapter 2, it seems implausible that this kind of pre-choice situation could indeed exist. However, I'd like to challenge the *conceptual impossibility* of such a choice and imagine a world where it could very well be that someone is really in two minds about which to choose: should they opt into agency or should they opt into shmagency?

I welcome the reader to treat this world as a hypothetical one. A world similar to ours, where a pre-agent is torn between the two possible options before them. In this particular pre-choice

situation, given reasons for being an agent weigh just as heavily as given reasons to be a shmagent. It is at this point that the individual wills a consideration to be a reason for her to opt into one rather than the other. Crucially, a voluntarist reason cannot overrule a given reason, but it can certainly take a given reason to be a voluntarist reason.

Perhaps one might worry here that what we take to be the consideration for our voluntarist reason might be arbitrary and so we cannot ground normativity there.

If I were torn between becoming a philosopher and a Trapeze artist and I had equal given reasons to choose either, I do not think that even if I were to take the smell of musty books as my consideration to choose to become a philosopher over a Trapeze artist, that this is an arbitrary consideration. Bearing in mind that I understand the depth and magnitude of this decision and that it bothers me very much to be in this situation where I have equal given reasons to choose either, the smell of musty books is not a randomly picked consideration, it's one that I take as my own and so is meaningful to me – it may mean nothing to you, perhaps to you the smell of musty books would be a reason to choose to be a trapeze artist instead – but this is not your decision and this is not your future. It's mine and I care about it very much.

But perhaps there's another problem here: Consider the circularity charge – how can an agent find reasons to be an agent outside of agency? Especially when having reasons is the business of agency. Maybe someone could respond that it would be impossible for an individual who is not yet an agent or a shmagent to have reasons in the first place (given or voluntarist) since perhaps having and generating reasons is the business of agency. But I can imagine an individual or entity who does not satisfy the criteria of agency who could have reasons. For example, a six-year-old child who, in the face of losing both parents to the prison system for

instance, takes on the responsibility of taking care of their younger sibling. A six-year-old is not an agent – but making that choice is an active buy-in into agency. Why?

Because the child, before this decision has no responsibilities and is not held accountable for any of her actions in any meaningful way. It's not the norm for a child to take on the responsibility of raising a younger sibling and it is not a burden an agent would push onto a child. Taking on this kind of responsibility is to acknowledge and accept that her actions now have a status in the world – a sense of accountability.

But perhaps this is too extreme an example. This certainly doesn't pertain to everyone. Notice though, that this objection misses the point. The incoherence and circularity charges pointed to the *conceptual* impossibility of choosing to be agents and having reasons from a place of non-agency. The child example shows that this impossibility claim is mistaken.

If we can accept this, then is this where the buck stops? Do we embrace agency and then we're agents from here on out? The answer is no, and this is something I argue for throughout this chapter. We choose agency often in our lifetimes – when we make the big decisions –such as whom to marry, or whether to immigrate, or whether to attend Cambridge or Oxford, we are embracing our agency.

### *5.6.2 Agency or Shmagency: Running out of Reasons*

I start with how it is we can run out of reasons to choose agency over shmagency. In § 4.3.3, I discussed a problem with Velleman's account. He pushes for the idea that the drive for sense-making is a mediator of our first-order desires etc and that this drive pushes us to do what makes most sense to do. However, if two things being equal are what makes most sense to do, how do we choose one over the other? In this way, Velleman's account is unable to show us



how to rank reasons. I also pointed out that in order to combat Enoch's insistence that we need to show why agency is an important thing to care about, the constitutivist would need to show why, and even how it is that we have reason to be an agent rather than a shmagent.

Because each account I discussed in this project commits wholeheartedly to agency being unavoidable, the ranking of reasons to be an agent rather than a shmagent would be inapplicable to them. The reason for this is if I asked any of these theorists why they would rank being an agent higher than being a shmagent, they would respond that agency is unavoidable so there's no actual ranking process that takes place. But, as I have argued for in § 4.4.5, the unavoidability of agency is not a virtue of these accounts as they'd like to think, it is just one more fundamental problem as to why these accounts fail. The way to move past this problem is to play around with the idea that perhaps agency is not as unavoidable as these theorists like to think.

For now, because this is a black box notion, I do not need to show you how choosing agency is possible. This I leave for the next chapter. Right now, let's revisit Julian, the accountant.

Because he has willed the consideration of feeling smart to be the reason he continues to be an accountant, he has created an expectation for himself to continue being an accountant. This now takes on a subjective value for him, he imbues the consideration with normativity – thus it becomes a reason for him to continue. What I want to take from this example is that it is possible to generate voluntarist reasons to choose to continue to do what one is currently doing. And also, that it is possible to discontinue doing what one is currently doing.

### *5.6.3 Choosing Agency*

When Enoch presents his challenge to constitutive accounts, he is asking them to provide reasons as to why we should be agents rather than shmagents. As I have already shown, a ready answer to this is that we have no choice as agency is unavoidable. But I've also shown how this is an unsatisfactory answer.

Another way to answer Enoch might be something like: 'okay, let's say agency is actually not unavoidable, let's say it's something we can opt into'. So then, we choose to be agents rather than shmagents. Maybe Enoch would then respond, 'why would you choose agency over shmagency, what is your reason for doing so? We have equal reason to do both. Why care to be one over the other?'

And that's a good question. We can use Railton's response and say that experiences and beliefs of a non-agent won't be yours and that's why you should choose agency. But then, Enoch might say, 'be that as it may, shexperiences and shbeliefs won't be yours as an agent and that's why you should choose shmagency'. This would seem like a reasonable challenge. What would probably happen in this situation, after long moments of deliberation, is that we would run out of given reasons to choose one over the other. Now, agency is a big deal. It's not something like choosing between two meals. So, we shouldn't just randomly pick one. What we can do instead is to will a voluntarist reason to be an agent over a shmagent. Maybe I take beliefs rather than shbeliefs to be a voluntarist reason to be an agent.

Maybe Enoch will then respond, 'okay, but then we can also will a voluntarist reason to be a shmagent rather than an agent'. And the answer is yes, you can. But then, you'd be a shmagent and you wouldn't care about adhering to norms because you'd have opted into adhering to shnorms instead. Most probably, you wouldn't be part of this conversation in the first place.

If we can generate voluntarist reasons to be an agent, then Enoch effectively has no leg to stand on. No longer can he demand of a constitutivist that even if what is constitutive of agency is to adhere to norms, or to accept norms, that we need to prove why we should care to be agents first. If we choose agency, then it follows that we care very deeply about being agents. Why? Because higher-level choice is both valuable and value-conferring. When we choose agency, we confer value onto agency – my agency now becomes valuable to me in virtue of my choosing it. Furthermore, because I will a consideration as a voluntarist reason for me to choose agency, I imbue this reason with normativity. I am a thing, which chooses and confers normativity.

What is the consequence of us choosing agency? So far, most of the accounts discussed in this project have tried to show how normativity serves as the bedrock of agency. However, if it is the bedrock of agency, and agency is something we have little to no choice over, then normativity is something we have little to no choice over. This is not ideal – mostly because we then run into objections such as Enoch's. A possible way to get around this is to entertain the idea that elective agency is actually something we can *choose*. If normativity serves as the bedrock of elective agency, and elective agency is something we can choose, then normativity is something we generate. What would this mean in the big picture? It would mean that we *should* care about norms because this too is something that we choose.

To further support this, there is empirical support for the claim that we care about those things we choose. For example, Henkel and Mather (2007) conducted a study where they monitored how people were likely to favour the option they chose over the option they didn't choose. According to Henkel and Mather (2007, p. 164) "the objective of a choice generally is to pick the best option. Thus, after making a choice, you are likely to harbor the belief that the chosen option was better than the options you rejected". Furthermore, "remembering the option that

you chose as being the best option should help reduce regret about your choice” (Henkel & Mather, 2007, p. 164). We would not experience regret nor believe that one choice is better over another if we didn’t *care* about what it is we choose.

Let me return to the example I used in the previous chapter of the mountain-climber. When the elective mountain-climber (this is the mountain climber who *chooses* to climb that mountain) stands at the base of Mount Everest and looks up, she knows that somewhere above her is a peak she wishes to stand upon. She understands that many have tried and failed. And she hopes that she will make it into the league of those that succeed. She has planned carefully for this. With her backpack secure, she begins the ascent. She chooses to do this. Her choice confers value onto being a mountain climber – one who will successfully climb this mountain.

Whenever she feels overwhelmed throughout the climb, she might pause and be compelled to ask: why am I doing this? She might provide enough reason as to why standing upon that peak is an end she wishes to attain – this will, or at least should, push her to move upward and onward. The mountain-climber continuously chooses to climb, to perform the actions necessary for her to achieve her end. Arguably, the best way to achieve her goal is to adhere to norms. To put one foot in front of the other, to remain focused, to use an ice-axe instead of a soccer ball to climb. To violate a norm in this setup could cost her her life, never mind be a bad way of attaining her goal. Imagine: mid-climb, she decides that she’d like to clip her toenails instead of using the ice-axe. The elective mountain climber is far more invested in climbing the mountain than someone who accidentally finds themselves climbing a mountain.

Let’s imagine the accidental mountain climber for a moment. This is the person who wakes up to find themselves strapped into a harness clinging to the edge of a cliff. They did not choose this and so did not confer value onto climbing this mountain. If anything, this poor accidental

mountain climber can't even feel regret which might come from making a bad choice, because she didn't make a choice at all! All she can feel is panic and maybe resentment for the person who put her there. But, when she chooses the ice-axe instead of a soccer ball, she confers value onto the ice-axe that she does not confer onto the soccer ball. If the ice-axe helps her to safety, then she won't regret choosing the ice-axe over the soccer ball.

Let's say also that she is able to use an ice-axe like a professional because she has an uncanny twitch in her left wrist, something she has *absolutely no control over*. She is deftly able to hook and pull in a way that only mountain climbers who have been doing this for decades are able to do. We can't say that she is deliberately adhering to ice-axe norms – but, she is definitely involuntarily doing it. If we ask her, why she is choosing to handle the ice-axe that way? She will embarrassingly answer that it's not her fault, this is just something happening to her. If we ask her, does she care to do it that way? She may answer: not really, but it seems to be getting the job done.

Because this is a black-box notion, I can be creative with the way I imagine how choosing agency happens. I imagine that choosing agency *does not* happen at a particular age or at a particular stage of development. I am less interested in how the initial choice occurred than I am in the continuation of that choice. I imagine that an individual chooses agency over and over again, on a daily basis. I choose it over and over again and so confer value onto agency as opposed to shmagency. My choosing agency is the highest-level of choice one could perform. And because this is the highest-level choosing, my choice is further value-conferring. I confer value onto being an agent.

When I choose agency, I also, as a consequence, confer value on that which is constitutive of agency. So, for instance, if rationality norms are constitutive of elective agency (something I

argue), then by electing agency, I also confer value on rationality norms – in particular, value which is subjectively important to me. Thus, accepting rationality norms is something worthwhile for me to be doing, in virtue of its being constitutive of agency. An elective mountain climber confers value on the norms associated with mountain climbing, for instance, being a mountain climber means adhering to norms of safety first as far as you are able to. This means choosing to purchase the best possible harness because you confer value on being a good mountain climber, because mountain climbing is something you choose.

One could try to argue that the unavoidable mountain-climber cares very much about *staying alive* and that's why she should care to climb the mountain well. But this is just a misconstrual of genuine norm-adherence. If anything, that sounds more like coercion than anything else. It's almost like putting a gun to someone's head and saying: 'you best climb that mountain well or else you'll die'. In any case, no one would agree that the person who adheres to the norms of mountain-climbing under these circumstances is doing so voluntarily and because they genuinely care.

This is just like an agent who finds herself being an agent and is unable to escape. How do we demand of this agent that they should care to adhere to norms which are constitutive of her agency? This agent is essentially in the same boat as the unwilling, unavoidable mountain-climber. Unless an agent chooses to be an agent, it is unclear how she is able to care about that which she chooses – incidentally, those things which are constitutive of agency – such as rationality norms, and it is further unclear why being an agent would matter in the first place.

#### 5.6.4 *Accepting Rationality Norms*

So, then the question is, if it is the case that we are able to elect to be agents, how is this then related to rationality norms? Well, I could argue for an instrumental connection between norms and agency and this can be fairly easy to do. I could say this:

Because agents are individuals who act, it makes sense that actions which are governed or guided by rationality norms are more likely to succeed than actions which do not. For example, imagine an individual who has been hospitalised for an appendix which is sure to burst, and she knows it. Let's call her Geraldine. Geraldine is writhing around in the hospital bed in agony. She meets with two doctors (separately). The first doctor enters the room and tells her that he will remove her appendix in the next hour and all will be fine. The second doctor enters the room and tells her he will remove her left arm in the next hour and all will be fine. For Geraldine to live through the next 24 hours, it is a very good idea for her to seek the treatment of the first doctor rather than the second – this would be the most rational thing for her to do.

But imagine a world where Geraldine's actions are not governed by rationality norms, imagine she is being governed by shnorms instead. I imagine shnorms look like norms, but are something different. Say Geraldine is keen to survive the next 24 hours and is approached by two doctors, the first offering to remove her appendix and the second offering to remove her left arm. It would be the most rational thing to do to get the appendix removed if her goal is to survive the next 24 hours. But, because rationality norms don't govern her actions, but shnorms do instead, she chooses, instead, to have her left arm amputated. As a result, Geraldine's appendix bursts and she dies.

So far, I have shown how to link norms and agency in an instrumental way. But the problem here is that I only show value is conferred on choosing one action over another insofar as they serve as means to an end. But by only showing an instrumental connection between norms and

agency, I fall victim to the eternal-sceptic Gary (Railton's student, discussed in Ch 3). Gary can ask why it is that he should care to adhere to norms. The instrumentalist can answer that he should care insofar as he wishes to attain an end. The problem is if all that links rationality norms to agency is that adhering to norms will help me be an agent, then why should I bother being an agent as opposed to a non-agent?

The above is an example of norms associated with agency in an instrumental sense. When Geraldine makes a decision (guided by rationality norms) to have her soon-to-burst appendix removed, she does so in an instrumental sense. The norm-guidance here is purely instrumental. A constitutive account would look at this quite differently. When Geraldine goes to hospital because she is suffering from a soon-to-burst appendix and she is approached by two different doctors, one offering to remove the appendix and the other offering to remove her arm – Geraldine chooses to remove the appendix. Why? Because Geraldine, in virtue of being an agent and choosing agency, accepts rationality norms. She is guided by them whether her decisions serve an instrumental outcome or not (this constitutive story is strengthened in chapter 7 where I show the mechanism behind elective agency).

Of course, at this tricky juncture, the constitutivist reminds us that we are agents already, and unavoidably so, and so asking the question seems rather meaningless. This answer is available to the constitutivist because they have pinned an *ultimate value* on agency. All things important, such as the source of normativity can be grounded in this ultimate value because it is apparently there for the taking. The best thing about it is that because of the unavoidability of agency (as other constitutivists would argue), everyone's got it. But, as we have seen, there are many issues associated with assuming the unavoidability of agency (something I talk about in the previous chapter).



It appears that constitutivists can only identify this ultimate value insofar as agency is unavoidable. I have argued already (especially in all the sections regarding norm-violation) that just because we *find* ourselves being something (like an agent), it doesn't follow that it is something we should care about. Remember also that if agency is something unavoidable and adhering to norms is just something that agents do, then are we ever truly adhering to norms rather than acting in accordance with norms? Norms which are inviolable are not norms.

At least, under my proposal, I identify value conferral in the truest sense. The way to truly establish value in a way that we can ground normativity in a constitutive account of agency is to argue for the following: by *choosing agency* day-after-day, perhaps minute-after-minute, I am endorsing rationality norms over and over again. The idea here is that *choosing* confers value. Under this new constitutive account of agency, agency is chosen and so, as a consequence, norm-acceptance is chosen too.

What makes my proposal more attractive than the other constitutivist accounts so far as that Gary, as well as Enoch, can no longer ask the why-care to be an agent question. As long as other constitutivists remain loyal to the notion of unavoidability, sceptics can ask this question.

Perhaps a worry at this point is that it may not be clear enough that elective agents do indeed accept rationality norms. But, given the *constitutive* nature of elective agency, rationality norms are part and parcel of embracing agency. Therefore, if you have chosen agency, you have chosen rationality norms. I have already said that we have equal given reasons to either choose agency or shmagency and so I cannot tell you why you chose agency over shmagency – you know what your voluntarist reason was.

#### 5.6.5 A Quick Recap

If accepting rationality norms is going to be constitutive of agency, as we are here assuming, then to answer the why-care question, we need to be able to show that adhering to norms is something we actually do genuinely care about. Second, we need to distinguish between unavoidable agency and elective agency and note that an unavoidable agent has no genuine vote in the acceptance of rationality norms. As this was something forced upon her, she cannot successfully explain why she should care to adhere to rationality norms as opposed to shnorms for instance. Third, because an elective agent chose, and continuously chooses to be an agent, she is in a far better place to answer the why-care question because accepting these rationality norms and being governed by these norms ensure that agency actually works out for her, thus validating her choice to be an agent.

So then, my argument looks like this:

P1 Accepting rationality norms is constitutive of high-level choice.<sup>70</sup>

P2. High-level choice is constitutive of elective agency.

C. Accepting rationality norms is constitutive of elective agency.<sup>71</sup>

In the next chapter, I explore the success rate of the black-box notion of elective agency with regards to the Shtest.

## 5.7 Conclusion

---

<sup>70</sup> Thanks to Asbjorn Steglich-Petersen who made me aware of the following issue: P1 precludes a natural worry: It appears that making any choice would mean that we accept rationality norms and so it seems an unnecessary move to invoke the move to elective agency. This is not how this works. It's not just any old choice I am talking about here. It's the choice of choosing elective agency which is tied into accepting rationality norms. As I have argued throughout this chapter, it's high-level choice I'm talking about, not arbitrary choices.

<sup>71</sup> Perhaps you are concerned that accepting rationality norms is constitutive of choice, then why do we need to make the additional move to agency? This is something I address in more detail in chapter 7 when I offer the mechanism behind the elective agent.

In this chapter, I introduced a new constitutive account called elective agency. By pushing the argument that agency is just something we fall into, other constitutive accounts push with it, the notion that rationality norms are also just those things we find ourselves accepting and again, something we are unable to escape. This has far-reaching consequences for questions about sustaining the normative grip of rationality norms and about why we should care to even be agents. The consequences are that the normative grip of rationality norms is an exceedingly weak one under this description and that there is no clear answer why we should care to be agents, it's just something we find ourselves doing.

The foundation of elective agency is that we choose to be agents rather than anything else. If we invest in this idea of elective agency, we are better able to meet the four problems as set out in the Shtest – better than any other account discussed so far (this I show in the next chapter).

The next chapter will examine how well the account of elective agency deals with the Shtest.



# Chapter 6 – Elective Agency and the Shtest

## 6.1 Introduction

In the previous chapter I introduced a new constitutive account of agency called Elective Agency, which serves as my contribution to the constitutivist debate as well as to debates concerning the source of epistemic rationality. This constitutive account is built on the foundational premise that agency is something we are able to choose rather than something we are thrown into, thus, challenging the claim that agency is unavoidable. This was done in the following way:

First, I positioned choice as that which is constitutive of voluntarist reasons. I did this by showing that at times, we run out of reasons and, through an act of willing (choice), we are able to generate voluntarist reasons (building on Chang). Furthermore, that we also choose given reasons to be those that serve our actions. Second, that we are able to endow reasons with normativity through choice, again, building on Chang's claim that when we create voluntarist reasons, we endow it with normativity. Third, that there is a constitutive link between choice and value-conferral, such that if we choose something, we confer value on it. Finally, that we choose agency and, as a consequence, we confer value on agency as well as endow it with normativity.

In this chapter, I analyse how well the account of elective agency fares with regards to the Shtest. The chapter is comprised of four sections. In section 6.2, I briefly set out the terms of the Shtest again as well as examine how the account of elective agency is best able to address the challenges set out by the Shtest. In section 6.3, I address possible objections to my account. Finally, I conclude in 6.4.

## 6.2 The Shtest: Pass or Fail?

In chapter 2, I discussed four problems that any constitutive account of agency needs to address in order to successfully pass the Shtest and hence to effectively answer the why-care question. Here, I convert those problems into design constraints on a constitutive account of agency which wants to successfully pass the Shtest. I argue that because we are no longer committed to the unavoidability of agency but can choose agency instead, we are in a better position to meet the design constraints set out in the Shtest.

I briefly go over the four problems here and show how the black-box notion of elective agency proposed in the previous chapter is able to address each. I argue that the account of elective agency comes closer than any other account in addressing the challenges set out in the Shtest.

### 6.2.1 *Show how Norm-violation is Possible*

How do I explain deliberate norm-violation as an elective agent? When I choose to be an agent, a choice I make over and over again (in some instances by taking a given reason or by willing a voluntarist reason to choose agency over non-agency or shmagency), I choose that which is constitutive of agency and that which is associated with agency. The distinction between my account and those I have already discussed in my thesis is that, where other theories see *adhering* to norms as constitutive of agency, I see *accepting* norms as constitutive of agency. In §2.5.5, I discussed several distinctions: the distinction between norm-awareness and norm-acceptance; the distinction between norm-adherence and norm-violation; the distinction between norm-adherence and agency; and finally the distinction between norm-adherence and norm-acceptance.

Norm-adherence, as I have already discussed, is to be aware of a norm and to following it. In the Thailand example, when I abstain from stepping on the coin because I know this to be a norm and I adhere to it in virtue of being a norm. Norm-adherence (in the other constitutive accounts I have explored so far) will have it that all we *can* do is obey the norm and thus all we *can* do is adhere to it as this is what is constitutive of agency – to *adhere* to norms. Norm-acceptance is different. Norm-acceptance is essential for norm-violation, and because of this, it is also essential for deliberate norm-adherence.

Think again of the Thailand example. If I do not accept the norm of not stepping on money, then it won't matter to me in the least if I do indeed step on money. Further, it wouldn't matter to me in the least if I abstain from stepping on money because this is a norm I have failed to accept.

One might object that a norm may not matter to me, but I do deliberately violate that norm anyways. For instance, a sociopath doesn't care for the norm to not hurt others, but still deliberately goes out to hurt others<sup>72</sup>. Two points in response.

First, the sociopath understands that this is a norm for others but does not accept that it should apply to him. He cares very little about it. He still proceeds to hurt others and by doing so is violating the norm of not harming others. Second, the sociopath is aware of the norm as something that *does* apply to him but considers it to be one that he actively does not care about. He proceeds to purposefully violate the norm. Note, that by accepting that this norm should apply to him and that, given different circumstances – such that if he were not a sociopath, he

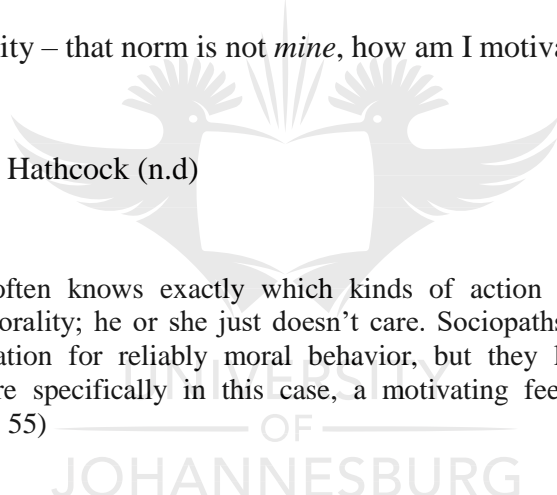
---

<sup>72</sup> Thank you to Veli Mitova for this example.

would have deliberately adhered to this norm – is enough of an acceptance of the norm to constitute deliberate norm-violation as I mean it here.

In the first instance, the sociopath has heard others speak of the norm that we should not harm others. Perhaps he understands how this can be important for someone else. But he actively does not buy into the idea that this norm is something he should care about. Perhaps his own satisfaction and happiness takes precedence over all others. He seeks out his victims thinking that the norm of not harming people apply to others, but not to him. Maybe he even rests assured that no one will harm him because of this norm. And so, even if the norm-violation looks genuine (and it is, to a large degree) it lacks a certain amount of ownership – and so too a lesser sense of self-accountability – that norm is not *mine*, how am I motivated to act in respect of it?

According to Johnson & Hathcock (n.d)



The sociopath often knows exactly which kinds of action are prescribed or proscribed by morality; he or she just doesn't care. Sociopaths may possess the cognitive foundation for reliably moral behavior, but they lack the affective foundation—more specifically in this case, a motivating feeling. (Johnson & Hathcock, n.d, p. 55)

Johnson and Hathcock (n.d) note that moral behaviour is dependent on two factors. First, moral knowledge which requires, at a minimum, that a person is able to identify a norm and know what kind of effects acting out this norm will have. Second, and more importantly, a person must be “sufficiently motivated that they perform a moral action” (Johnson & Hathcock, n.d, p. 55). In the first instance then, the sociopath has the necessary knowledge (the cognitive foundation) but lacks the motivational drive (affective foundation). Both are needed for deliberate norm-violation.

In the second instance, the sociopath has heard others speak of the norm that we should not harm others. He actively buys into the idea that this is something which should apply to him.

Perhaps he recognizes himself as an active member of society and the community and because of this, accepts that he should adhere to the norm of not harming others. But this does not automatically translate into him adhering to this norm. He accepts that this norm should apply to him but does not accept it as a norm he will adhere to at this time. This is different from the first instance in an important way. Here, the sociopath has both the cognitive foundation and the affective foundation in place. This is an instance of deliberate norm-violation then.

When I accept a norm, I understand that this is a norm which I endorse, I choose this norm as one of my own and so, I confer value onto it. When I choose to purposefully violate this norm, I am deliberately undermining the choice for this to be a norm of mine in the first place. Perhaps, you could say, that through choosing to violate the norm, I am conferring value onto the *violating* of that norm – a kind of negative-value conferral if you will. Perhaps this is true. In the next chapter, I discuss how it is possible for choice to also be value-diminishing in some instances, not necessarily primarily value-conferring.

Recall my overarching argument:

P1. Accepting rationality norms is constitutive of high-level choice.

P2. High-level choice is constitutive of elective agency.

C. Accepting rationality norms is constitutive of elective agency.

When I choose to be an agent, I am choosing that which is constitutive of agency: accepting rationality norms.<sup>73</sup> Because accepting rationality norms does not immediately equate to norm-

---

<sup>73</sup> Note a possible objection that to value an entity does not entail valuing all its elements. So, just because I might value agency, it does not mean I automatically value all that is associated with it. I deal with objection under §6.3.5



adherence, but can equally result in norm-violation, the elective agent is able to do either. Because of this, my account of elective agency is in a better position to show how norm-violation is possible.

Here, I'd like to make use of Geraldine again to illustrate my point. In the previous chapter, Geraldine made the right choice and had her appendix removed – she is now thankfully happy and healthy. Suppose now that she is training for the marathon. It's 5:30 in the morning and she is running along the edge of a lake. Geraldine stops when she hears a cry for help. She sees a person drowning in the middle of the lake. Now imagine the following scenario.

It's 5:30 in the morning and Geraldine is running alongside the lake. She hears someone crying for help. She pauses to deliberate what would be the best thing to do. In the past, if anyone had ever asked her what she would do in this situation, she would have easily answered that the best thing to do would be to jump into the lake to save the drowning person. However, now she cannot move toward the water because, instead of thinking about jumping in, she is thinking that she really wants a cigarette and a cup of coffee instead. She decides to go for the cigarette and coffee. When she looks out over the water again, she cannot see or hear anyone. She immediately runs home and informs the police.

In this scenario, Geraldine set a goal to save the drowning person, but instead wants a cigarette and a cup of coffee instead. There was no fear hindering Geraldine's ability to adhere to norms. She was fully capable of doing so. Here, deliberate norm-violation has taken place. – importantly, in both instances, Geraldine's agency is neither diminished nor eradicated in these acts of norm-violation. When she chooses agency, she accepts rationality norms. When I accept rationality norms, I can either mistakenly or deliberately violate norms. Note that this is

different from the constitutive accounts that state that agents who adhere to norms in virtue of being agents, sacrificing agency (to an extent) when violating norms.

Let me explain this further. The account of elective agency is not saying that because an agent violated a norm, she is any less of an agent. Instead, for the elective agent, it diminishes the value of continuing to be an agent. Ideally, an agent should not be that thing which considers having a cup of coffee and a cigarette rather than saving a life – there's something unattractive about being this type of agent. Because the elective agent chooses to be an agent, she chooses to accept rationality norms. When she commits to wanting to save a life and chooses to have a cup of coffee and a cigarette instead, she is violating a set of norms she chose (and continues to choose) to accept. But furthermore, she is violating the worth of being an agent. This calls into question her decision to choose agency in the first place. If she cannot adhere to those things she accepts, should she be playing this game of agency? These kinds of actions diminish the reason for her to choose agency over non-agency. Thus, norm-violation challenges her choice to be an agent, without making her lose agency.

### *6.2.2 Do not Appeal to Unavoidability to Answer the Why-care Question*

To answer the why-care question, a constitutive account of agency should not answer that we can't but care in virtue of being agents. Just because agency is something we cannot avoid – at least to the constitutivists discussed so far – it doesn't mean that this is an automatic reason as to why we *should* want to play the agency game over the non-agency game. Further, this is not reason enough to care to adhere to norms.

The way elective agency avoids this problem should be obvious: we should care to be agents because we choose to care – in other words – because we confer value on that which we choose, not because we have no other option. If an elective agent knows that she continuously chooses

agency over non-agency, then she is seeking to be the kind of agent which will confirm that this choice as a good one (I discuss this in the next chapter). Second, because the elective agent chooses to be an agent, choosing norms that guide her actions is something she cares about in virtue of choosing them. Choosing to be an agent puts her in a unique position to care about her agency as opposed to finding herself being an agent, involuntarily.

To return to the mountain-climber example. The unavoidable mountain-climber, who happens to find herself strapped into a harness, clinging to a mountain-side, cannot tell us why she should care to adhere to mountain-climbing norms without appealing to the fact that she has to adhere to the norms to survive this ordeal. And, importantly, this is purely an instrumental consideration – the kind of which will not be sufficient to put Enoch’s concerns to bed. However, if survival is not at the top of her list, then really, adhering to mountain-climbing norms is as important to her as adhering to mountain-climbing norms. But all this has very little to do with her caring about *mountain-climbing*. So, when the unavoidable mountain-climber claims that she must care because she is unavoidably a mountain-climber, we can see why this is not a satisfactory answer because it doesn’t mean she automatically cares about mountain-climbing in virtue of being an unavoidable mountain climber.

The elective mountain-climber, in contrast, is a different story. The elective mountain-climber is able to show she chooses mountain-climbing because *mountain-climbing* is an important thing for her to be doing. She confers value onto it insofar as she chooses it. She can wholeheartedly commit to caring to adhere to mountain-climbing norms because she cares to be a successful mountain-climber. Just as she can wholeheartedly commit to caring to adhere to mountain-climbing norms because she may care to be a successful mountain-climber. Furthermore, because accepting mountain-climbing norms is constitutive of being an elective mountain climber, and because this is something of her choosing, to be asked why she doesn’t

adhere to mountain-climbing shnorms instead is to essentially call into question her choosing mountain-climbing over shmountain-climbing. The elective agent can simply respond that shnorms are not what she agreed to accept when she chooses to be an agent.

And, maybe Gary, the eternal sceptic could ask of me why it is I choose mountain-climbing over shmountain-climbing, if both options are on the table. As discussed in the previous chapter (§5.5), if we have equal reason to choose either, we are in the unique position to will a voluntarist reason to serve as my reason to choose agency rather than shmagent. Because choice is value conferring, when I choose agency, I confer value onto agency. Furthermore, I choose those features which are constitutive of agency – such as accepting rationality norms.

### *6.2.3 Show how Features are Non-arbitrary*

A constitutive account of agency needs to show how the features which have been identified as constitutive are non-arbitrary, essentially showing how they are different from other desires or other features. Also, to show how these constitutive features are imbued, somehow, with a normative-uniqueness, or a privileged normative-status. Choosing agency is what confers value on agency. As has been argued previously, if we have any hope of grounding the source of rationality norms, our best bet is to do so in high-level choice because through high-level choosing we confer value. This is what allows the constitutive feature to hold its privileged normative-status. There is an ultimate value in high-level choosing in which rationality norms can be grounded and, as I previously argued, choosing agency is one of the highest levels of choice where we can imbue normativity. As it happens, choice is the hallmark feature of the elective agent and because of this, rationality norms can be grounded in elective agency. High-level choice is thus non-arbitrary.

Non-arbitrariness can be further established if we establish that other things do not share this constitutive feature – that it is truly unique to agency. If you remember one of the concerns with Korsgaard’s walking example (see chapter 3) is that the constitutive aim of walking can also be shared by running and jumping. Furthermore, I may have desires, such as taking a nap under a tree, which are similar to the desires my cat might have. So then the desire to take a nap under a tree is arbitrary, just as the constitutive feature of walking is. High-level choice is different. It’s difficult to point to anything else but an agent and say it has high-level choosing capabilities. Even if we consider a raven or a chimpanzee and prove its incredible intelligence – I don’t believe it is there that you will find high-level choice, not the kind that is the constitutive feature of agency.

The non-arbitrariness of high-level choice, however, becomes far clearer in the next chapter where I discuss the mechanism behind how this choosing is enabled.

#### *6.2.4 Don’t claim Self-vindication*

A constitutive account, when answering the why-care question, should not say that by asking that question, we are using the very feature of agency whose significance is being questioned. This is a constraint which all the other accounts have failed to meet and I explained in the previous chapter that the failure can be attributed to the close link constitutivists forge between the unavoidability of agency and self-vindication (§ 4.4.5) This constraint requires, of a constitutive account of agency, that the answer not lie in a self-vindicating property which is the constitutive feature of agency. Here, I argue, there is actually no way to get away from the problem of self-vindication, regardless of the account of agency you choose to defend. I further argue however, that the account of elective agency deals with it best because elective agency does *not* see self-vindication as a problem.

Perhaps the concern here is, if I don't see self-vindication as a problem for elective agency, why did I saddle the other accounts with this particular burden? The simplest answer is that Enoch is the first to do this – and remember the Shtest is largely modelled on the constraints he sets out, avoidance of self-vindication being a prominent one. Furthermore, all the constitutivists I have discussed so far run up against the problem. I argue below that even Enoch's own alternative – robust metanormative realism – also runs up against the problem.

I've also argued in chapter 4 and 5 that self-vindication is certainly a problem if paired with the problem of the unavoidability of agency. Something all the constitutivists as well as Enoch himself does. The reason self-vindication is not a problem for the account of elective agency is because this account does not pair self-vindication with unavoidability. And so, this is what makes the account of elective agency a more attractive account. In §4.4.5, I argued that the three accounts fail because they too intimately connect the unavoidability of agency and the problem of self-vindication insofar as these accounts derive self-vindication from the unavoidability of agency, they will always find themselves in a sticky situation. This is not to say that if this intimate link were not there, the problem would be entirely solved. However, the problem of self-vindication becomes far less of a problem for the constitutivist if we break this intimate link between the two.

What this link does, however, is to not allow constitutivists much elbow room. It doesn't allow for one to identify a feature which is non-arbitrary or unique. Instead, it forces constitutivists to answer the why-care question by pointing to the very thing asking the question and saying, well, as far as this thing is unavoidable, this is what you are asking this question with, ... or, this is the position from whence you ask.

This is clearly unacceptable. But it is not a problem for the elective agent. There are elements of the enterprise of agency which are as any other enterprise. If I were to agree whole-heartedly with Enoch that agency is an enterprise *no different* from another, I am shooting myself in the foot because of the constraint regarding norm-uniqueness and norm-arbitrariness. Let me explain. Agency cannot be an enterprise as any other – like chess – because then it doesn't have the non-arbitrariness needed to ground the normative grip of rationality norms.

But, if agency is a vastly different enterprise from all others, then we also find ourselves in hot water. If agency is an enterprise which we just happen to find ourselves in, then, and I have argued for this already, the norms are those things which we just adhere to. This makes it difficult to show how agency and norm-adherence are something we get to choose – which is the most important feature of elective agency. The challenge then, is to show how the enterprise of agency is similar to other enterprises without foregoing the norm-uniqueness and non-arbitrariness requirement.

But surely this challenge demands a sacrifice somewhere in the account. The enterprise of playing chess is a game we decide to play with the aim of achieving a checkmate and because we decide this, we voluntarily accept the norms of playing chess. The answer to why we should care to follow the norms of playing chess is because we chose to play the game and thus chose to accept the norms constitutive of chess-playing. We would fail to be playing chess if we chose to play the game of chess but played by the rules found in monopoly instead. In this way, the norms of chess are unique to the game of chess and hold some normative grip over the chess-player in virtue of her having chosen chess in the first place. But it doesn't have enough normative grip to keep the chess-player from haphazardly forsaking the game of chess to play

Monopoly instead. Even if the norms of chess are unique to the game of chess, it is a weak normative<sup>74</sup> hold it has over the chess-player.

This type of weak normative hold will not work with agency and this is perhaps what makes agency less of an ordinary enterprise than chess-playing. When it comes to playing the game of elective agency, we do so by accepting rationality norms. Because we chose to play this game, we should care to follow rationality norms of playing the game. But the difference between the enterprise of chess-playing and the enterprise of agency is that agency requires more loyalty from its 'player' than chess does. The features constitutive of agency are thus, essentially different from chess-playing. Chess-playing has an aim to achieve checkmate and not a lot rides on achieving this checkmate other than satisfying a desire to play chess well. There is more to agency than a mere desire to play the game well. Agency is far more valuable an enterprise.

I have argued that choice is unique to the enterprise of agency and that choice confers value. And so, to ensure norm-uniqueness, a constitutive account of agency would need to show how norms which are either embedded in these features, or arise from these features, are those which we are bound to. In other words, that these features have a normative grip in a way that no other features can. Arguably, with elective agency, the norm-uniqueness is embedded in the ultimate choice of choosing agency. This highest-level choice is very different from other levels of choice as has already been discussed.

Note however, it is possible that there are elements of agency which are similar to any other enterprise, such as the ability to choose to play the game, and thus choose to be guided by these norms associated with this game. But it is further possible that the enterprise of agency is

---

<sup>74</sup> When I say a weak normative hold, this is not here used technically and is not related to weak norm-violation.



distinct from other enterprises which allows for the normative grip of rationality norms – more so than a typical enterprise would have. A way to do this is to show how the constitutive features are such that confer value for instance – something I have already argued for in the previous chapter and something I explore in more detail in the next. But, does this get us away from the problem of self-vindication?

I believe it does, *only insofar as we break that intimate connection between self-vindication and the unavoidability of agency*. I discussed (in §4.4.5) how, if we break this link, we don't have to view self-vindication as a serious problem anymore. And this is yet another aspect which makes elective agency a more successful account than other constitutive accounts discussed so far. That even if we concede that we cannot ask for reasons to play the game of agency without actually already playing the agency game, we at least choose to be in this position to play the agency game in the first place – thus, effectively breaking the link between self-vindication and the unavoidability of agency.

Think for a moment of asking a philosopher why it is that they think being a philosopher is more worthwhile than being a trapeze artist. When the philosopher responds to you that discussing the meaning of life is a far more worthwhile intellectual pursuit than swinging through the air under the big top, she is actively engaging in the process of philosophizing when providing you with this answer. This is where self-vindication rears its ugly head. There just seems no way that the philosopher can answer that question without philosophizing – philosophizing being a distinguishing constitutive feature of being a philosopher.

The unavoidable philosopher (and I'm sure some exist), cannot engage with the world through a different, non-philosopher lens. They are compelled to ask and answer the questions through the process of philosophizing. This, I believe, would make for a difficult life. Always seeking

out the deeper meaning in having a cup of tea or analysing the nature of bread before eating your lunch does not sound particularly fun to me and, worse, it is an unavoidable process. There is just no way for our philosopher to escape philosophizing, however dull or challenging it might be at times. Even in justifying this dullness or defending it, the philosopher must do so through the process of philosophizing. This is very different from the elective philosopher.

The elective philosopher can opt out of seeing the world through a philosopher-lens any time they wish, by choosing to be a non-philosopher instead. When ordering a cup of tea, they may switch into non-philosopher mode and not care that its tea or coffee either way. And perhaps, yes, if we ask the elective philosopher why being a philosopher is more worthwhile than being a trapeze artist, they would have to use the very constitutive feature of being a philosopher to do so – philosophizing. But the distinction here is that the elective philosopher willingly opts into this role, whereas the unavoidable philosopher does not.

What does this mean for the self-vindication constraint then?

It means that when we ask the elective philosopher why it is more worthwhile to be a philosopher rather than a trapeze artist, they do indeed use the very constitutive feature of being a philosopher to answer the question – they philosophize. Admittedly, this is an instance of self-vindication. The concern, as pointed out by Rosati (and Enoch, and Velleman, and even Railton) is that

their operation [the features constitutive of agency], we might say, is self-vindicating, and efforts to challenge them cannot even get going without relying on them. (Rosati, 2003, p. 522)

But, and this is where the problem may essentially lie

the very challenge presupposes, it seems, that it can be met, because the activity of evaluating these motives and capacities – presumably an instance of one of the activities agency consist in – already presupposes them and their significance. Take

away the motives and capacities constitutive of agency and you have taken away agency itself and with it the challenge to the normative significance of these motives and capacities. (Enoch, 2006, p. 45)

The reason we presuppose that the challenge can be met is because we presume that the motives and capacities (those things which are constitutive of agency) are already, and automatically so, in place in virtue of the fact that an agent is the one posing the challenge. Furthermore, we presuppose the significance of these features in virtue of the fact that they are constitutive of agency.

Two things the account of elective agency offers here. First, we can't presuppose the challenge can be met until we are sure agency has been chosen over shamagency or non-agency. There needs to be an active buy-in to agency for this presupposition to work and this isn't always clearly the case.

Second, Enoch argues that if we take away the constitutive feature of agency, which I argue is choice, then we have taken away agency "and with it the challenge to the normative significance of [choice]" (Enoch, 2006, p. 45). But this seems a bit strange. Because as much as Enoch is pushing for the notion that the source of normativity cannot be grounded in the constitutive feature of agency, there seems to be a push for grounding it in something else – what he calls robust metanormative Realism. Enoch argues that when we remove the constitutive feature of agency, we remove the normative significance of that thing. I argue that this worry is not mutually exclusive and occurs under both the constitutive account as well as the robust metanormative account.

Enoch proposes a robust metanormative realism as an alternative. I think this is likely going to face the same negative consequence he has set out for the constitutivist, especially regarding this constraint. Robust metanormative realism seems to hold that just as scientists have had to

depend on the existence of the electron to explain how their theories or experiments work – thus appealing to the inference to the best explanation –

Similarly, the fact that our best attempts at deciding what to do and how to live our lives require normative facts (indeed, irreducible normative ones) gives us just as good a reason to believe in normative facts. [...Furthermore] that the deliberative project is indeed nonoptional in the relevant sense; that irreducibly normative truths are indeed deliberatively indispensable. (Enoch, 2006, p. 57)

It just seems to be the case that if we are to deliberate why we should be agents, we are bound to appeal to this nonoptional deliberation to do so, thus presupposing the capacity and its significance. Arguably, because there is no constitutive link, we cannot presuppose that just because we take away the nonoptional deliberation, that we take away agency – but then the challenge to the normative significance of this capacity still holds. We have an agent with no capacity for nonoptional deliberation, but which still needs to, somehow decide “what to do and how to live our lives” (Enoch, 2006, p. 57). The normative significance of this capacity is further pronounced when our agent fails miserably at making decisions in virtue of the agent not having this capacity. Nonoptional deliberation is instrumentally necessary.

Enoch’s alternative to the constitutive account is strange in two ways. First, Enoch commits to the notion of the unavoidability in a different sense when he speaks of this deliberation as being nonoptional (agents are either those things which engage in this nonoptional deliberation or are those things who fail agency in virtue of not engaging in this nonoptional deliberation). Second, he sets up a purely instrumental notion of the source of normativity which has already been shown to be problematic (see the eternal sceptic Gary’s concerns).

How is the elective constitutivist better able to deal with the constraint of self-vindication than any of the other constitutivists or the robust metanormative realist? The unavoidable philosopher has no other option, in virtue of being a philosopher, but to philosophize when

answering why being a philosopher is a worthwhile thing to be doing. The robust philosopher (the philosopher under the robust metanormative realism theory Enoch proposes) – who is also an unavoidable one –has to philosophize when answering why being a philosopher is a worthwhile thing to be doing in virtue of it being the best way to answer the question. The elective philosopher philosophizes when answering why being a philosopher worthwhile thing to be doing in virtue of choosing to be in this position. Self-vindication occurs in all three scenarios.

Maybe there is no clean way to get around the self-vindication constraint, but perhaps there is the *best* way to deal with it. Perhaps choosing to self-vindicate seems a better way to cope with the self-vindication constraint than to say we have no option but to self-vindicate. But this doesn't seem right – self-vindication is self-vindication regardless of whether you were thrown into it or choose it. Maybe, and this is something that is more likely to be the case, the self-vindication problem is actually not a problem at all, and so is therefore no longer a constraint. How is the elective constitutivist to show this?

We can look at this in two ways:

First, I have already argued that you cannot derive norm-uniqueness or norm-significance from the constitutive features of the unavoidable agent – because there is no clear account why I would adhere to the norm just because this is something one does in virtue of being an unavoidable agent. And so, when we adhere to rationality norms in answering the question (we are using the constitutive feature to do so in virtue of being an unavoidable agent), then by answering the question, it is not vindicating the norm-significance of these rationality norms (the constitutive feature) the way constitutivists would have you believe – because we cannot establish norm-uniqueness from an account that pushes the unavoidability of agency.

Second, I have argued that you can derive norm-uniqueness from the constitutive features of the elective agent because there is a clear account of why it is I care to adhere to norms in virtue of choosing it, thus conferring value on it. Then, by adhering to rationality norms when answering the question, we are vindicating the constitutive feature as well as the norm-uniqueness of that feature.

### 6.3 Objections

I now consider eight possible objections by way of strengthening my proposal.

#### 6.3.1 *First Choice?*

The first possible objection I would like to consider is one which I imagine Gary, the eternal sceptic would raise. Let's say that Gary buys into this idea that agency is something we choose on a daily-basis, that by choosing agency, we are conferring value on it. But because Gary is a student who enjoys pushing boundaries, maybe he will ask the following question 'if we are things which are continuously *choosing*, how did the first choice happen?'

My response is this: I am not bound to the notion of a fixed beginning. I am only committed to the notion of continued choice. Of course, there must have been a first choice to be an agent, it would be implausible for me to say there wasn't one. But I don't think it matters much how this happened. What does matter is that we continue to choose to be an agent and that we have the choice to deliberately opt out if we so wished.

Examples of opting out of agency could include getting very drunk, taking drugs, or even sleeping. Importantly, Railton (2003) speaks of opting out of agency in a far starker sense. He considers opting out of agency as something as severe as suicide. I don't think this is

necessarily the case. There are plenty of times in our average lives where we purposefully forego agency – thus opting in and out of agency. When one chooses to undergo anaesthesia for surgery – this is an active and deliberate choice to opt out of agency. When one gets hypnotized to maybe stop smoking, one opts out of agency. When a person decides to drink far too much alcohol, one opts out of agency, and these are very temporary.

Arguably, it might be difficult for someone to opt back into agency from within these states. I certainly won't be in the position to opt into agency whilst lying on a surgery table under the effects of anaesthesia. But that doesn't mean I don't opt into agency once the anaesthesia wears off. Of course, there are a few conditions that arguably have to be in place in order for one to choose agency – at a minimum, consciousness might be a basic requirement. Think of opting in and out of agency as analogous to the access one might have to a little island called *Juist*, off the coast of Aurich, Germany. If you wish to access the island by foot, you can only do so when the tide has gone out. This allows you a clear walk to the little island. Of course, when the tide rises, there's no way to walk back. You have to wait for the tide to go out again in order to return to the mainland. Just so, you can opt out of agency and you can opt back in when the conditions are right.

In chapter 5, I provided the example of Julian, the accountant who escapes to his car during lunch to deliberate whether he should continue being an accountant or switch his career to that of ping-pong playing. In this instance, Julian can opt out of being an accountant if there is enough reason to do so. Just such is the case of agency. When we are faced with equal reason to choose either agency or shmagency, we can will voluntarist reasons to choose agency, and this is essentially what we do, that's why we're agents rather than shmagents. Again, I am not bound to a fixed beginning of choice, but rather to *continued-choice*.

But, perhaps it could be objected that I cannot equate choosing agency to Julian choosing to continue to be an accountant. Julian may choose to continue to choose to be an accountant because he has to do something. A pre-agent does not have to do anything. There is no urgency to choose one thing over another nor is there any desire to do so from a pre-agent's perspective. But again, this is assuming a world where a pre-agent is not a subject and is therefore unable to have any features that an agent or a shmagent might have.

There are features which are not unique to agency which other things can also have – such as the desire for milk for example. I can have the desire to add milk to my coffee and my cat can desire to have milk in his saucer. I can desire to have more responsibility at work for example (perhaps I will get a promotion then) and a very young child can have a desire for more responsibility when it comes to chores (perhaps it means more pocket money). An agent may a desire to have a home and a squirrel may desire to have the same thing.

Having to do something is not unique to agency. An urgency to choose between two options is also not unique to agency and therefore it's possible that a pre-agent has this need to do something just as an agent does.

### *6.3.2 Why be an Elective Agent?*

A second possible objection could again be raised by the eternal sceptic Gary, and potentially Enoch: 'okay, but even if elective agency is something we choose, why should we choose this rather than elective shmagency?'

I feel that I have dealt with this sufficiently in §5.5.1. But I'll go over it briefly here again. When it comes to deliberating over whether to choose agency or shmagency, perhaps you have given reason to choose shmagency and if you have chosen shmagency, then you willed this to



be the case. However, it also means that you probably wouldn't be reading this right now. The fact is that when you deliberated over which to choose, you probably ran out of reasons. And, following on Chang, you willed a voluntarist reason to choose agency over shmagency. Your reading of this is evidence of this choice.

Of course, what is problematic here is that Enoch has not stipulated very clearly what a shmagent might look like or what shnorms look like. (And I have followed him in this vagueness.) So, it's entirely possible that a shmagent could read this as well as an agent. Presumably, a shmagent is not the direct opposite of an agent (that's the role of a non-agent). Presumably, there may be infinite possibilities of groups of norms ranging from similar to agency to ones that are completely different from agency.

For example, perhaps shmagency<sup>1</sup> has the same set of norms as agency except for one, perhaps that these shmagents never modify their beliefs when facing contradictions.<sup>75</sup> The next type of shmagent, call it shmagent<sup>2</sup> is shmagent<sup>1</sup> minus the rules of inductive inference. This could carry on to shmagent<sup>100</sup> who is perhaps least like an agent. The problem then is that it's entirely possible that shmagent<sup>1</sup> could be reading this as well as an agent.

The problem with this kind of characterisation of shmagency and agency is that it buys into the narrative of an all-or-nothing agent. Let me explain. Korsgaard's agent forgoes agency when violating a norm. As well as the other constitutive accounts which view norm-violation as something an agent probably just does not do or does so by mistake. According to the characterisation of the shmagent above, when Korsgaard's agent upholds a belief in the face of contradictory evidence, she slips from agency into shmagency, not non-agency. For Velleman's agent, when they uphold belief in the face of contradictory evidence, it's what

---

<sup>75</sup> Thanks to Veli Mitova for pointing out that this is problematic.

makes most sense for them to do as an agent. They don't slip from agency to shmagency, but rather from agency to another form of agency.

I'm going to bite the bullet here and say that a shmagent is most likely that thing which upholds shnorms which are almost all different from an agent's norms. I'm further going to say that an agent can appear shmagent-like when she violates her norms, but this does not mean she is a shmagent. It means she is an agent that violates norms. Maybe you'll want to challenge this by saying 'it's not about the norm-violation as such, it's about whether that thing holds the norm to begin with'.

A shmagent does not accept the norm: change your beliefs in the face of contradiction – so adherence to this norm is certainly not expected of a shmagent. We find ourselves in a weird position then. What happens if shmagent<sup>1</sup> violates the *shnorm* of not changing beliefs in the face of contradictory evidence? This would mean that she is now adhering to the *norm* of changing beliefs in the face of contradictory evidence, incidentally the *only* shnorm/norm that was separating a shmagent from an agent. Does the shmagent<sup>1</sup> now become an agent? If this is true, is she *opting* into agency if she knowingly and voluntarily violates this shnorm, but accepts this norm? Given this puzzle, I think we need to think of shmagents as being very different from agents.

Furthermore, *ought* we to choose agency over shmagency? This is difficult for me to answer because I can't point to a given reason and say, that's the reason you're an agent rather than a shmagent. I can't point to a given reason because, remember, we have given reasons to choose either. Therefore, the voluntarist reason you chose to be a reason is one I may not be privy too. Suffice to say, it's possible that in the process of genuinely being on the fence, you bring into existence a voluntarist reason which may be very different from mine.

### 6.3.3 *The Elective Agent Cannot Choose to not Accept Rules*

Because I am arguing for a constitutive account of agency and because I posit accepting rationality norms as constitutive of agency, then it just doesn't seem plausible that I can ignore what is constitutive of agency and still remain an agent. In other words, it shouldn't be possible for me to be able to deliberately violate a norm, not in the way I have shown in §6.2.1, and still be an agent. It's the same as saying that a roof constitutes a house, that I don't care to use a roof, but I still want to build a house. Just because I can choose to build the house, does not mean I get to choose not to use a roof.

My response is twofold. First, what is constitutive of agency is accepting rationality norms, not *adhering* to them. So, when I want to build a house, I accept that I can use a roof to do so, but I am not forced to use a roof. Granted, it might not make for the strongest house if I choose to use something else instead of a roof, but building a house is still something I am doing. In the process of choosing agency, I could also reassess that which is constitutive of agency, including the acceptance of rationality norms – as the *acceptance* of norms is what is constitutive of agency – not the adherence to norms. Violating norms does not make me less of an agent, it allows me to reassess the norms which I have committed and am committing to.

Second, because I sever the link between the unavoidability of agency and self-vindication, it doesn't matter if, in the process of caring about norms I am falling back on the constitutive feature of agency to do so. Self-vindication is only problematic if the very feature which is constitutive of agency is one I happen to find myself with and one which I cannot change. This is not the case because of electiveness.

### 6.3.4 *Do We Really need a Constitutive Account of Agency?*

A further possible objection is to ask why it is, if accepting rationality norms is constitutive of choice, that we need to then link it to agency. Why not stop the buck at choice, rather than latching it to agency?

My response is that even though accepting rationality norms is constitutive of choice, choice and action have a constitutive aim: to always seek out worth in being an agent, in seeking out validation for the choice to continue to be an agent. This becomes far clearer in the next chapter where we discuss the mechanism behind elective agency.

Without this constitutive aim, choice on its own could be perceived as rudderless.

#### *6.3.5 When Choosing, it does not Automatically Mean I want the By-products*

This objection deals with the idea that just because elective agency is chosen, it doesn't mean that things that are constitutive of it are chosen.<sup>76</sup>

For example, when a person chooses to take a shower at the gym, you are choosing to clean yourself after a sweaty workout. You are further choosing to perhaps wash your hair and shave your legs. But suppose you contract athlete's foot. Your intentions when using the gym shower was so that you could clean yourself. Not once did you step into the shower and silently pray 'I really do hope I get athlete's foot. That's why I'm here'. But even though your intention was to clean yourself, you also managed to take on something else that was not of your choosing.

---

<sup>76</sup> Thank you to Veli Mitova for pointing this out.

Maybe someone could say that choosing agency seems to be a bit similar. One could say that they chose agency and ended up with a bunch of unwelcomed norms they were somehow forced to accept in virtue of choosing agency.

My response to this objection is that it rests on a disanalogy. Luckily, athlete's foot is not essential to, let alone constitutive of, shower-ness, while in the same way that accepting norms is essential to the agency. I have argued that value is conferred insofar as our actions are chosen under the guidance of rationality norms. The idea is that value-conferring stems from an unconditional value, for instance, rationality choice. And so, it just isn't possible that through our choosing agency, we are in the position to not choose accepting rationality norms. The two go hand-in-hand.

#### 6.3.6 *Is Elective Agency not also Inescapable?*<sup>77</sup>

Could it not be objected that elective agency itself falls back on us making a choice to be an elective agent, rather than making a shmoice? In other words, is it not the case that we are already engaging in the business of agency when we are talking about making choices rather than being in the business of shmagency where instead we would talk about shmoices?

My response is two-fold. First, I'm not sure what making a shmoice would look like and if it's even necessary to make a shmoice to become a shmagent rather than make a choice to become a shmagent – if we begin to ask for shmreasons to shbe shmagents and shmoices that will get us there, we are beginning to ask questions which border on being intelligible. I want to avoid this as far as possible. Second, if we can imagine that choice can also belong to those things which don't necessary qualify as an agent (in other words, doubting the exclusively of choice

---

<sup>77</sup> Thanks to Christoph Hanisch who provided for this objection.

as belonging only to agents, as I argued in §5.6.1), then I can imagine I can make a choice to become a shmagent as much as the choice to be an agent.

### 6.3.7 *The Bleakness Objection*<sup>78</sup>

Another objection might be that the account of elective agency may come across as unpalatable and rather bleak.<sup>79</sup> There is something bleak about thinking that agency was chosen when we ran out of reasons. One might even think that agency, a deep-seated fact about every individual, is established upon an *arbitrary* decision when reasons have run out. One wonders if it isn't best to be condemned to agency in the way Korsgaard means it, compared to a form of self-damnation where I ran out of reasons and arbitrarily chose agency and now have to spend my life vindicating this choice.

I ask the reader at this point to please note that elective agency is absolutely not derived from some random, arbitrary choice. Just because non-voluntarist reasons may have run out, does not mean we randomly pick just anything to be a consideration for my choosing agency. Note, that choosing agency is a high-level choice and you cannot treat this kind of decision as you would between two meals (chicken or fish). It's okay to plump for a reason between chicken and fish, but absolutely not okay to do this with a high-level choice such as agency.

Voluntarist reasons do not amount to random, arbitrary reasons and to think they do is disanalogous. The process of realising one's given reasons have run out to taking a consideration (and note that the consideration can still be a given reason) as a voluntarist reason

---

<sup>78</sup> Many thanks to Jessica Lerm for making me aware of this potential problem.

<sup>79</sup> Thanks to Jessica Lerm for this objection.

is a *big deal* and one that is done with investment from the individual themselves. I don't see this as arbitrary, I see it as empowering in a way that other constitutive accounts do not allow.

Because of this, being condemned to agency, as Korsgaard would have it, seems far more unpalatable to me than to choose agency through taking a consideration as a voluntarist reason for me. In the world of elective agency, agents have a sense of buy-in in their agency rather than having no other choice but to accept it agency as something they arbitrarily fell into.

### 6.3.8 *The Phenomenological Objection*<sup>80</sup>

Another possible objection to the account of elective agency is that it seems phenomenologically jarring. If we do indeed choose agency, then we are wholly phenomenologically unaware of it. It's difficult to point to a particular time and say, ah yes, that's the moment I chose agency because that's when I started to feel like an agent.

I have a two-pronged reply to this. First, it must be noted that having a sense of agency is already phenomenologically *thin*. We are not phenomenologically aware of our agency as we are aware of our vision for instance, or of our sense of touch. Certain conscious experience is phenomenologically strong, unfortunately however, the feeling of agency is not. Therefore, the *sense* of agency is already something we find difficult to quantify. Second, because agency is phenomenologically thin, and because you do not know what it feels like to have agency, how can you know if you have ever chosen agency in the past or if it is something you still do? If you are unable to tell me what agency feels like, then you would be unable to know that agency is something you did not choose.

---

<sup>80</sup> Thanks to Jessica Lerm for this.

Notice, though, that this can go both ways. Perhaps someone may object that if it is so phenomenologically thin, how do you know you can *choose* it? Proponents of viewing the sense of agency as a combination of the ‘model of apparent mental causation’<sup>81</sup> (Wegner, 2002) and ‘feedback-comparator model of motor control’<sup>82</sup> (Haggard, 2005) differentiate between a basic sense of agency and reflectiveness concerning one’s agency.

The basic sense of agency is construed as an online and phenomenologically rather thin experience that accompanies the performance of actions [...and] Judgments about one’s agency, in contrast, are offline and usually post-act, and they are, thereby, subject to various biases that may distort the interpretation of one’s own agency. (Schlosser, 2019).

Think of the six-year-old pre-agent who decides to take on the responsibility of looking after her younger sibling in the absence of parents – in the act of choosing to do this, the six-year-old experiences the basic sense of agency, note again, that this is phenomenologically thin – but that does not mean it is phenomenologically missing entirely. But, it’s phenomenologically thin enough for us to not be able to state that ‘ah yes, that’s the moment I chose agency because that’s when I started to feel like an agent’.

## 6.4 Conclusion

In the previous chapters, I argued that most constitutive accounts of agency discussed so far commit to the unavoidability of agency and this amounts to quite a few failures in the Shtest. Thus, unable to sustain the normative grip of rationality norms. I argued that the way to get around this is to show how it is that we choose agency rather than to be thrown into it.

---

<sup>81</sup> According to this view, our sense of agency is brought about when we actively interpret our intention to perform a particular action.

<sup>82</sup> “According to this model, the motor control system uses copies of motor commands in order to generate predictions of the ensuing bodily movements. Those predictions (so called “forward models”) are then used for comparisons between the predicted and the intended trajectories of movements, and for comparisons between the predicted and actual trajectories (based on information from sensory feedback)” (Schlosser, 2019).



In this chapter, I provided the results of this account of elective agency. I put it to the Shtest. I set up the challenges set out in the Shtest as design constraints and argue that if something is to pass the test, it must be able to meet all four design constraints.

I showed that my account is able to meet the first design constraint: showing how deliberate norm-violation is possible. I argued that when we choose agency, we also choose that which is associated with agency – or in this instance, what is constitutive of agency.

Second, I showed how my account is able to meet the second constraint: do not appeal to the unavoidability of agency in answering the why-care question. My account is best able to meet this constraint for an obvious reason, we choose agency, it's not something we fall into, and it thus does not suffer the problems associated with unavoidability. How does it answer the why-care question? Because we confer value through choosing, we care about that which is constitutive of agency – such as accepting rationality norms.

Third, I show how my account is able to meet the third constraint: to show how what is the constitutive feature of agency is non-arbitrary. If high-level choice is what confers high-level value, then choosing agency is at this highest level that we imbue normativity. There is nothing arbitrary about this in the least.

Fourth, I show how my account is able to best meet the fourth constraint: to not claim self-vindication when answering the why-care question. I argued that insofar as self-vindication is derived from the unavoidability of agency, self-vindication will always remain problematic. By breaking the intimate link between the notion of unavoidability and self-vindication, self-vindication becomes far less of a problem for the elective constitutivist.

In the next chapter, I explore the *how* behind this black box notion of elective agency. What I do is propose, similarly to Velleman's, a constitutive drive, but different from Velleman's in that this drive is one that aims at establishing worth. Here, I show the constitutive nature between accepting rationality norms and the worth-drive – thus further explaining how it is that we should care to adhere to norms. I also show, more conclusively that choice is the hallmark feature of agency and that accepting rationality norms is something one does in virtue of embracing the worth-drive. I further show that the worth-drive is the constitutive feature of agency. Thus, establishing the constitutive link between accepting rationality norms (through embracing the worth-drive) and agency.



# Chapter 7 – The Worth-Drive

## 7.1 Introduction

In the last two chapters I offered a new constitutive account I called Elective Agency. What is unique about this account is that it overcomes the problems set out in the Shtest in a way that the other accounts could not.

The elective agent is one who has the capacity to choose agency and because of this, she chooses to accept rationality norms which are part and parcel of agency. The elective agent is able to do better at answering the why-care question because, in her choice to be an agent, she is committed to caring to be an agent, and as a result, is committed to caring about that which is constitutive of agency, such as accepting rationality norms. Thus, an account of elective agency is better able sustain the normativity of rationality norms.

One thing I did not focus on in the last chapter was *what* elective agency is. That is the focus of this chapter. I provide here a substantive account of elective agency is in terms of what I call the worth-drive.

This chapter is comprised of seven sections. In section 7.2, I introduce the worth-drive and explain how it is that thing which we choose when we choose to be elective agents. In section 7.3, I examine Frankfurt's theory that action is value conferring. I extract what I need from this account to explain what I mean by the drive for worth, or the worth-drive. In section 7.4, I show how the worth-drive essentially works by showing how actions are either in a position to validate or undermine the decision to be an elective agent. In section 7.5, I argue that it is in virtue of the worth-drive that this account is more successful with regards to the Shtest. In section 7.6, I strengthen the proposal by addressing possible objections.

## 7.2 Just What *is* the Elective Agent?

An elective agent<sup>83</sup> is different from unavoidable agents in virtue of agency being something that can be chosen, or, in some instances, rejected. Here, I argue for following:

- i.* There is a drive which aims at establishing worth. I call this the worth-drive.
- ii.* Accepting rationality norms is constitutive of the worth-drive.
- iii.* The worth-drive is constitutive of agency.
- iv.* When we choose agency, we embrace the worth-drive.

In this section, I briefly discuss each of these in turn. (I explain them in more detail throughout this chapter).

### 7.2.1 *What is this Drive?*

A drive, in the sense that I mean it here, is a functional state which motivates or guides an individual to think and behave in a certain way. There are many different kinds of drives out there, notably the more obvious ones, such as a sex-drive or, in Nietzschean fashion, the drive-for-truth (Nietzsche, 1988).<sup>84</sup> Just as the sex-drive predisposes one to seek out partners with which to procreate, a drive for truth may be a predisposition to always seek out absolute truth “beyond human life” (Nietzsche, 1988, p. 42). An important feature of a drive is that it is not merely reactive.

In other words, a drive does not merely wait about for a suitable stimulus in order to get activated, but actively seeks for discharge. Drives, therefore, manifest themselves as urging forces. (Riccardi, 2018)

---

<sup>83</sup> It is important to note however that there may be potential grey zones where we may not have complete certainty that something is indeed an agent. This is also true however of any important gradable concept.

<sup>84</sup> I use Nietzsche here to provide an intuitive understanding of what I mean when I talk about a drive. I do not intend on engaging with Nietzschean literature as other theorists, such as Katsafanas has done.

We have already encountered two drives in this project – Velleman’s drive for sense-making and Katsafanas’s drive for power. The drive that I propose here is a third one – a drive which aims at worth. This drive seeks to identify and establish value in objects, actions, persons and events. Something is of value or valuable as far as it is desirable or good. Some things may be more desirable or good than other things; and thus, some things have more value than other things. In the previous chapter, I explained that through high-level *choosing*, we confer value.

Value can depend on a number of factors. For instance, if we go back to the voluntary mountain-climber example. The ice-axe is of great value to our mountain-climber in virtue of potentially being that which will most likely get her to the top of mount Everest if she chooses to use the ice-axe. After the mountain-climber has achieved this successful ascent and descent and is now having a hot cup of coffee at home, the ice-axe now holds a different kind of value. For one, she is not dependent on it to achieve an end anymore and so the axe holds very little instrumental value. Although, it may be invaluable non-instrumentally as it represents the most important climb of her climbing career (I discuss this distinction between instrumental and non-instrumental value a bit later on).

The worth-drive aims at identifying and establishing value/ worth. I use these two terms interchangeably throughout. The worth-drive is a functional motivational state which drives an agent to pursue actions, make decisions and choices in a way which confers value on objects, events, and ultimately, agency. The idea then is that when we continuously choose to be an agent, we are continuously embracing this worth-drive. In the previous chapter, I argued that choosing, especially the bigger choices in life, is where one can ground the source of rationality norms. Here, I show how rationality norms are constitutive of the worth-drive.

### *7.2.2 Accepting Rationality Norms is Constitutive of the Worth-drive*

In §1.1, I explained, albeit briefly, what I mean when I talk about rationality norms. Here, I outline it again. When it comes to identifying objects, actions, persons and events as valuable, I cannot imagine an instance where this is achievable without the guidance of rationality norms. For example, an important principle of rationality is that we should not hold two contradictory preferences for essentially the same thing (Broome, 2003; Brunero & Kolodny, 2013; Davidson, 1982; Wallace, 2001). For instance, I should not simultaneously prefer (with the same intensity) having a cup of tea and not having a cup of tea. This kind of torn preference will result in inaction because my desire to have a cup of tea and to not have a cup of tea renders me unable to reach a decision about tea. The mountain-climber should not, if her desire is to reach the peak, prefer to use the ice-axe and not use the ice-axe at the same time. They should not essentially hold the same value for the agent.

The idea is that through embracing the mountain-climbing-drive (imagine there was such a thing), an individual chooses to be an elective mountain-climber. The mountain-climbing-drive motivates a person to confer value in the pursuit of being a good mountain-climber (or, in some instances, being a mountain climber at all), such as using an ice-axe, rather than a soccer ball, to scale up a wall. The drive ensures that the mountain-climber does not get into a situation where she values using an ice-axe and not using an ice-axe to climb the wall because this kind of contradictory belief would be to the detriment of the climber.

Another principle of rationality is that if an agent identifies that a particular action will bring about a particular end or goal – the agent should perform this action (Brunero & Kolodny, 2013). I only know when certain actions are valuable in comparison to other actions insofar as my value-assessment is governed by rationality norms and insofar as once I choose one, I confer value on it *for me*. It is likely the case that I will only know that *this* action as opposed to *that* is more valuable because this action will help me achieve my goal as opposed to that

action.<sup>85</sup> In Chapter 5, I argued that there is a difference between choosing and picking something at random. Choosing requires an ability to weigh up options to reach a decision – and if one runs out of reasons to choose one over the other, one is able to will a voluntarist reason to choose one over the other.

A further principle of rationality is that we should not hold contradictory beliefs (Broome, 2003; Wallace, 2001). If the mountain-climber believes (correctly) that the ice-axe in her left hand is made of carbon fibre, but also holds a belief (incorrectly) that the ice-axe is made out of chocolate mousse, she will either climb the mountain with it or attempt to eat it. In either case, the ice-axe holds vastly different values. One, as a carbon fibre ice-axe, the value is far higher if the mountain-climber wishes to reach the peak. Two, the chocolate mousse ice-axe is valuable if the mountain-climber should get hungry – but only if this belief were correct. The overall harm of holding either contradictory or false beliefs should be clear, if the mountain-climber attempts to eat the carbon fibre ice-axe, she will break her teeth. Clearly though, values are more clearly and definitively categorised when under the guidance and governance of rationality norms.

But, the above has *only* pointed to the instrumental value of accepting rationality norms. I argued in the Chapter 5 that something as important as grounding the source of rationality norms is to be found at a much higher level. I argue that the normative grip of rationality norms can be situated in *choice*. More specifically, the high-level *choice to be an agent*.

If choosing is value-conferring, then ultimate value can be found through choosing something as fundamental and unique as agency. But for this to work, there has to be something more

---

<sup>85</sup> This example is of course pointing to instrumental value, which is not the only value I am interested in here. In this chapter I address both instrumental and non-instrumental value.

concrete than a black box notion of agency to embrace. Something which will ensure we are motivated to act out that which is value-conferring, something as solid as a *functional motivational drive*.

I propose that when we embrace the worth-drive, we are embracing this high-level value. The worth-drive pushes us to act in a way that validates our choice to be an agent. This is a constitutive claim. If the worth-drive was not in the position to push us towards validating this choice, then it is less likely that our actions would be value-conferring, such that we would be in a less likely position to want to choose to be an agent as opposed to something else – such as a shmagent.

Think for a moment of our elective mountain-climber. The argument would go that if she had not embraced the mountain-climbing drive but happened to find herself on the edge of a cliff, then she is an involuntary mountain-climber. Adhering to norms in this instance is only instrumentally important to her. But let's think of another scenario. The elective mountain-climber is scaling up the mountain very happily. She is using all the correct tools which will get her to the top. She values herself as a mountain-climber as she climbs. Her actions are confirming that her choice to climb the mountain was a good one. The mountain-climbing drive is ensuring this.

But, let's say that unfortunately she slips and falls five meters and lands on an outcrop. She hits her head. When she comes to, she is able to comprehend that she is on an outcrop on the side of a mountain wearing a harness and has an ice-axe conveniently attached to her belt. She has also, in the process of hitting her head, foregone the mountain-climbing drive. Because of this, she has no desperate desire to reach the peak and feels a moment of sheer desperation because she wants off this mountain at all costs. She has now slipped into the role of the involuntary



mountain-climber. Without the drive present, those actions which are supposed to confer value on her choosing to be a mountain climber no longer do this. Those actions are one of survival – instrumental –and mean very little to her as an involuntary mountain climber.

The role of the worth-drive is to push us towards validating this choice to be an agent. This is possible because accepting rationality norms is constitutive of the worth-drive. Every action I perform as an agent is constitutively aimed at establishing worth. When we determine what is the most worthwhile action to take or the most worthwhile decision to make, we inherently accept and adhere to rationality norms when we do so.

What does this look like? For example, when I am deliberating if I would be best suited to be a trapeze artist or a philosopher, I am guided by rationality norms. Which one of these careers would best validate my choice to be an agent? This requires, at a minimum, that I do not hold two contradictory beliefs: I cannot believe that I want to be a philosopher and also that I do not want to be a philosopher. If I did hold these contradictory beliefs, my worth-drive would be very much frustrated – there is no way to validate my decision to be an agent if I am unable to make a decision because I hold two contradictory beliefs.

The constitutive link between rationality norms and the worth-drive is this: the worth-drive aims at worth, at both seeking and conferring it. Actions which are going to validate the decision to choose to be an agent the most are the ones that we are typically driven to choose. Thus, my actions are those which are most worthwhile to do. At a minimum, for an agent to determine which action is the most worthwhile one to do, she needs to adhere to rationality norms, particularly *consistency* (Mitova, 2016) and *reliability* (see §4.3).<sup>86</sup>

---

<sup>86</sup> Please see the discussion in §4.3 where I make the same claims regarding rationality norms and Velleman's drive for sense-making.

Imagine if Julian, the accountant, held inconsistent and unreliable beliefs. Perhaps he believes that to be an accountant is all about numbers and also not about numbers at all. Perhaps he likes the way numbers make him feel smart. But because he holds these inconsistent beliefs, he is unable to even will a voluntarist reason that accountancy is for him, because he can't make the connection between accountancy and numbers, hence between the choice of accountancy and the consequence of feeling smart in future. An accountant, for him, is all about numbers, and simultaneously, not about numbers at all. The worth-drive, in this instance, is unable to achieve its constitutive aim – it cannot establish worth.

### *7.2.3 Agents are Those Things which Embrace the Worth-drive*

Objects, actions, persons and events are valuable and hold a certain objective worth (this list is not exhaustive). But also, there is a subjective value which an agent confers through choice. Some things are valuable in and of themselves and others are valuable insofar as they further a goal or end of an individual. To argue that an individual who is unable to recognise the objective value in objects, actions, persons and events is still an agent would be a difficult thing to do. If agents are those things which act, it is essential that she is able to establish value through her choices – with regards to the goals she pursues – and just generally those things which add more value to her life. First, goals have value. This value can be non-instrumental (e.g., the value of finding love) or instrumental (the value of earning money).<sup>87</sup> To have non-instrumental value, the object must be valuable (desirable) in and of itself, regardless for what it can do for you or regardless of its functionality.<sup>88</sup> Means to achieve ends have (instrumental)

---

<sup>87</sup> For more on the distinction between instrumental and non-instrumental value, please see Korsgaard (1983).

<sup>88</sup> Some theorists refer to this type of value as final value. For instance, G.E Moore (1903) distinguishes something as having intrinsic value insofar as its value depends on the internal properties it has. Something has final value, in contrast, insofar as it has value for its own sake. Most philosophers use 'intrinsic value' to mean final value. However, as has been argued by some theorists, such as John O'Neill (1992) and Shelley Kagan (1998), it is entirely possible that final values are not always intrinsic. To avoid confusion, I will refer to an object which is valuable as an end as being non-instrumentally valuable.

value. Objects, actions, and sometimes persons which will better bring about the ends have value. But finally, and most importantly, persons, including the agent herself, are themselves bearers of value.

An elective agent is one who *continuously* elects to be an agent over a non-agent or, as has been dealt with in §5.5.2, a shmagent. The only way this is really possible is if she is able to establish the worth of being an agent as opposed to a non-agent. And, as a result of electing agency, is also able to identify and establish the worth of objects, actions, persons and events as these things are related to her. The ability to establish this worth, lies in the worth-drive. The elective agent is an individual who establishes worth for herself through her choices, particularly aiming to establish worth in being an agent. If I make such a big choice – such as that of choosing agency – I would want that choice to be confirmed as a good one. I don't want to have to regret choosing agency.

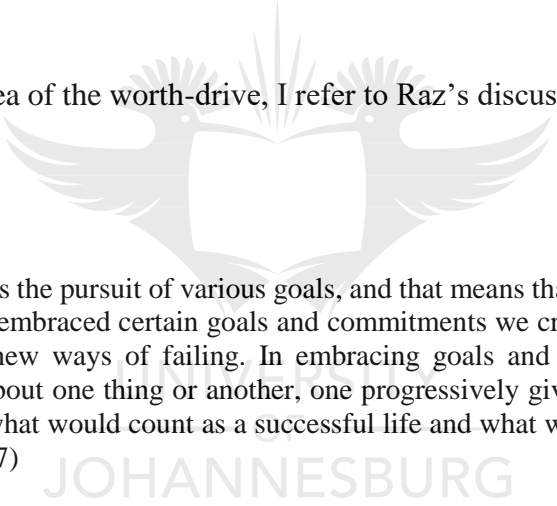
Imagine I choose to become a trapeze artist. Every night I climb up the ladder in my sequined leotard and chalk up my hands to swing under the big top. But, as I swing, every night, I am utterly *miserable*. My hands hurt, I get motion sickness, and often I find myself wanting to let go just so that it can all be over. I thoroughly regret the choice to become a trapeze artist. Every action I perform as a trapeze artist adds more and more to my misery and my regret. The bigger the callouses on my hands become because of the swing, the more I curse the day I chose this profession. This was a bad decision to make and every action I perform in virtue of this one decision adds to my regret for making that decision in the first place.

Now, imagine one day I have had enough, and I quit the circus. I decide to become a philosopher. As I read David Lewis's work for the first time, I am enthralled. When I publish my first paper, I feel immense happiness. When I give my first class and see the students

listening intently, I am delighted. I find myself thinking back to the moment I decided to be a philosopher and I *know* I made the right decision. I do not regret that decision in the least. I validate that decision every time I perform actions as a worthwhile philosopher.

How is it that we embrace the worth-drive when we choose agency? When I generate voluntarist reasons to choose agency over shmagency, I confer value on agency. In other words, I establish a metric of worth that I need to uphold in order for this decision to count as a good one. The only way to ensure that this value is recognised and carried out, is through the worth-drive. I need something which will guide my actions and decision-making and ensure that my choice to be an agent is a good one. Without the worth-drive, our choices are *rudderless*.

To further support the idea of the worth-drive, I refer to Raz's discussion of *self-creation*. Raz claims that



Our life comprises the pursuit of various goals, and that means that it is sensitive to our past. Having embraced certain goals and commitments we create new ways of succeeding and new ways of failing. In embracing goals and commitments, in coming to care about one thing or another, one progressively gives shape to one's life, determines what would count as a successful life and what would be a failure. (Raz, 1988, p. 387)

He goes on to say that

one creates values, generates, through one's developing commitments and pursuits. In that way a person's life is (in part) of his own making. It is a normative creation, a creation of new values and reasons. [...] The fact that one embraced goals and pursuits and has come to care about certain relationships and projects is a change not in the physical or mental circumstance in which one finds oneself, but in one's normative situation. It is the creation of one's life through the creation of reasons. (Raz, 1988, p. 387)

When we create voluntarist reasons to choose agency, we are essentially embracing a commitment to be an agent which reinforces this choice. And so, we are able to determine what “would count as a successful life” insofar as we have the worth-drive to do this. If we perceive

of the choice to be an elective agent as a commitment, we can better understand that we can only fulfil this commitment in virtue of embracing the worth-drive. When we choose, we commit, and so we embrace the worth-drive.

This I explain a bit more in the next sub-section.

#### *7.2.4 We Are Elective Agents in Virtue of the Worth-drive*

A couple of things need emphasis here. First, that an elective agent is one who continuously chooses to be an agent rather than a non-agent or shmagent. Second, that an elective agent, through accepting rationality norms, is one who *continuously* evaluates her reasons for, and her actions of, being an elective agent.

The more an action, decision or deliberation is able to validate the choice – where ultimate value lies – the more we have (voluntarist, subjective) *reason* to pursue that action, decision or deliberation. Remember that there are instances where she may have equal reason to either be an agent or a shmagent, and in some instances, she wills a voluntarist reason to choose agency. However, there are also instances where the reasons don't add up, in other words, there's more reason to choose non-agency or shmagency over agency and the elective agent has the ability to step outside of agency, foregoing all that is associated with agency. This can be a temporary or permanent decision.

This is vastly different from the kind of unavoidable agent as explored by other theorists. The difference is two-fold. First, the elective agent is able to seek for subjective reasons to be an agent from *within* the place of agency. An example I used in the previous chapter was Julian, the accountant who wills to continue to be an accountant over the option to be a ping pong

player. Second, the elective agent is able to step outside of agency whenever she chooses. Let me show you how.

In §6.2.1, I provided the scenario of Geraldine who takes a run along a lake and discovers that someone is in the process of drowning. After taking some time to deliberate which action to take, Geraldine is too late to save the drowning victim.

This incident plagues Geraldine and she is unable to shake off the feeling that she is no longer worthy of being an agent. For Geraldine, choosing to be an agent is also to accept certain norms, where adherence is optimal, and violation is optional. Some violations are more acceptable than others: safely crossing a street when the light is clearly red is OK, allowing a person to drown is not OK. *Is she worthy of being an agent if this is the kind of thing she does? Is this the kind of agent she chooses to be?*

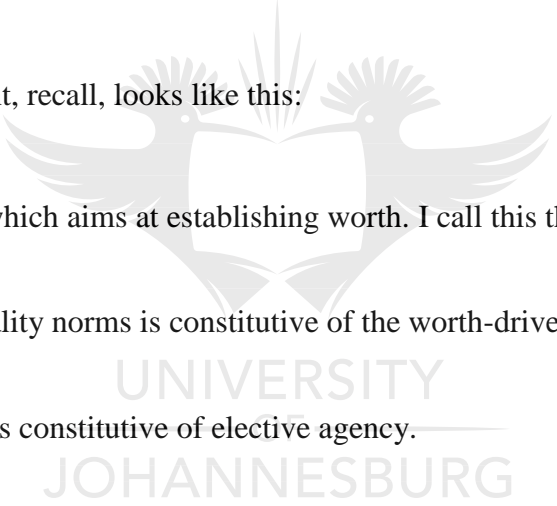
Geraldine has it in mind of what being an agent *means* and also confers *value* on being an agent – thus justifying her choosing it over and over. But she also knows that if she had not been an agent, she would not have been in the position to allow someone to drown. In these instances, she is unable to justify the choice to be an agent and begins to wonder if deliberating non-agency is something she should be entertaining. Nor would she feel the remorse and regret she is now feeling. In actual fact, it would be far better for her if these feelings did not belong to *her* but to a non-agent instead (thus acting on symptoms of an unfulfilled worth-drive). Furthermore, if this is what it means for her to be an agent and this is the kind of value she confers (in this instance, disvalue), then being an agent just seems to be less of an attractive thing to be doing. Why be an agent if this is the kind of thing she does?

To think that allowing a person to drown is an inaction *in virtue* of being an agent – then being an agent just doesn't seem all that worth it because not only did she negate the value of the

drowning person, but her own value has been diminished through this action. Perhaps, this realisation drives our agent to temporarily forsake agency. She goes mad and is institutionalised. Perhaps she is driven to drink excessively so that she never has to approach life sober again. Perhaps she opts for a more final termination – suicide.

To sum up. An agent's actions need to positively reflect this decision to be an agent as opposed to a non-agent or shmagent. If enough of her actions invalidate this decision often enough, this individual may be driven to question whether she should continue being an agent. Her actions thus, should be - and typically are - reflective of her decision to be an agent. And so, we choose agency by embracing the worth-drive. An elective agent is one who embraces the worth-drive.

My overarching argument, recall, looks like this:

- 
- i. There is a drive which aims at establishing worth. I call this the worth-drive.
  - ii. Accepting rationality norms is constitutive of the worth-drive.
  - iii. The worth-drive is constitutive of elective agency.
  - iv. When we choose elective agency, we embrace the worth-drive.

In this section, I explicated the notion of an elective agent by showing the following things. First, we are guided by the worth-drive. This drive aims to establish value. Second, to accept rationality norms is constitutive of this worth-drive. I showed how something which is not governed by rationality norms is unable to establish value. Third, I showed how agents are those things which embrace the worth-drive and fourth, that this is the hallmark feature of what it means to choose agency – to be an elective agent. In the next section, I go in more detail regarding what I mean when I say that the worth-drive aims at establishing value.

### 7.3 Value in the Worth-Drive

My claim is that through embracing the worth-drive, one chooses to be an agent. When we choose to be a shmagent for instance, we don't embrace the worth-drive. There are three things that we do when we make the decision to be agents. First, we *confer* value on agency by *choosing* agency. Second, we establish by choosing, that there is something *worthwhile* about *being* an agent. We further will a voluntarist reason to choose agency over shmagency, thus imbuing this decision with normativity. Third, acting in a way that adds more value to agency reinforces the decision to continue choosing to be an agent.

In this section, I explain some foundational concepts such as the distinction between instrumental and non-instrumental value. From there, I examine Harry Frankfurt's model of value. Here, I clarify the sense in which I use the worth-drive.

#### 7.3.1 Distinction between Instrumental and Non-instrumental Value

Value is often defined in terms of desirability or *goodness* (Foot, 1985; Nagel, 1985; Schroeder, 2016). An action or an object, relationship, or event, can be thought of as valuable if it is a good thing to do, have, or be in. Arguably, if an action holds little to no value, a person is far less likely to undertake the action in the first place. Potentially, if an action is a good one or will lead to good outcomes, it is considered good, by that person, or of value to that person. This is the point when things get a little tricky for this notion of value. If value is defined in terms of goodness, then we need criteria for what is considered good as opposed to bad or non-good.<sup>89</sup>

---

<sup>89</sup> These criteria can be comparative and gradable. Something is good, better and best in the way that something is short, shorter and shortest. Moral philosophers however, "write as if they assume, that 'good' is basic" (Schroeder, 2016, p. 9). Conversely, Consequentialists may argue that certain states of affairs are arguably better



Something has value insofar as it is desirable or good. Some things are more desirable than others; some things have more value than others. Value can be based on a number of things: utility, goodness, and preference. Value can be obvious, such as a Porsche being more valuable than a Honda because of a number of factors: scarcity of a product, monetary cost and perhaps demand. Value can be less obvious, such as when a person refuses to sell or give away an old, chipped tea set, or a baby blanket. A knife can be valuable insofar as it cuts well, or spreads butter well. A blunt knife is less valuable than a sharp one if the purpose of the knife is to cut well. A sharp knife is less valuable than a blunt one if the purpose of the knife is to spread butter well. If we have absolutely no use for the knife, then it is difficult to determine if it has any value at all. Unless it is an extraordinarily beautiful knife which is a pleasure to just look at, or if it is a knife from the Yuan Dynasty and has played a very important role in Mongolian history.

The Yuan knife can be valuable in two ways: it has both *instrumental* value and *non-instrumental* value. For example, say you were a proud owner of Yuan knife and its home is on top of your mantel piece. Everybody who visits your home comments on how beautiful this knife is and how lucky you are to have it. Other than the compliments you receive, arguably this knife has non-instrumental value given its history.

This knife is valuable as an end – its value is not dependent on what it can do for you or what it may do for you in the future – its value is non-instrumental to what it essentially is. Now, let's say one night you fall asleep on the couch and at three in the morning you hear a window break. You wake up in a panic and grab the Yuan knife to protect yourself against the intruder. The intruder sees you with the weapon and immediately escapes – terrified for his or her life.

---

than others and that a state of affairs is “good or bad in *virtue* of their being more things of value in them” (Schroeder, 2016, p. 14).

The moment you held the Yuan knife in your hand and had the intention of using it as a weapon, is the moment it took on instrumental value for you.

Just as we can recognise the Yuan knife as being instrumentally valuable insofar as it can serve as means to an end, we can also recognise agents as instrumentally valuable even though we should not recognise them for this alone.<sup>90</sup> Being an elective agent is valuable in and of itself, independent of what an agent can do for others. Of course, agents have instrumental worth insofar as they contribute to the population, economy, education, politics, social advancement etc. But agents should not be recognised as *just* those things which hold instrumental value. There is something non-instrumentally valuable about agency and about *being* an agent. When we choose to be agents, we are conferring non-instrumental, rather than instrumental, value.

When an agent is to determine the best possible action which will validate the choice to be an agent, she needs to understand the relationship between practical reason and value. There needs to be a clear relationship between the reason an agent has for acting and which action would be good for her to do.

I have already argued that high-level choice holds ultimate value (above and in the previous chapter). Through choice we confer value on things.

What we establish as valuable feeds into our reasons for action. We value agency and we value the actions of those of agents (rather than, say, shmagents or non-agents). We have less reason to perform actions which are going to diminish our choice to be agents and more reason to

---

<sup>90</sup> When you order a coffee from a barista, you recognise him as that person who is either going to make you an incredible cup of coffee or a very bad one. You depend on him to provide you with a cup of coffee which will hopefully make you happy. The barista is instrumentally valuable to you. Should the barista suddenly be unable to do his job, he becomes far less valuable to you. But the barista is non-instrumentally valuable in the grand scheme of things in virtue of being an agent. Agency is an end in and of itself and therefore is non-instrumentally valuable. The barista, as an agent, is someone we should recognise as having non-instrumental value, regardless of what it is this agent can, or cannot do, for us.

perform actions which will validate or confirm our decision to be an agent over a non-agent or shmagent. But how do we know which reasons are going to validate or diminish our choice to be agents?

We are in a better position to determine reasons for actions only if we can establish a strong account of rationality norms which is able to show how one feels motivated to act as one should rather than just as one *involuntarily does*. We should act in a way that does not diminish our choice to be an agent but endorses it. There is a very strong motivational pull to ensure this happens if we can buy into the idea that agency is something we choose by embracing the worth-drive. And, as has been argued before, this high-level choice is where one can ground the source of rationality norms.

In the next sub-section, I look at how some theorists treat action as value-conferring. From this discussion I hope to show how action, when it comes to elective agency, can be both value-conferring as well as potentially, value-diminishing.

### 7.3.2 *How does Value-Conferral Work?*

For my account of elective agency to explain the normative grip of rationality norms, there needs to be a clear and strong link between rationality norms and the worth-drive. I believe that the notion of action as value-conferring will get me there. The idea that action is value-conferring goes as far back as Kant. For the Kantian, a normative reason to perform an action is that it is a *good* thing to do – thus marrying practical reason and value. This kind of treatment

of action as value-conferring is explored, even if only tacitly so, by many constructivists, such as Korsgaard.<sup>91</sup>

In this subsection I discuss Frankfurt's notion that action is value-conferring. It is not my intention to bind myself to a specific view or theory of value, but rather to show that regardless of how we may think of value, the worth-drive is constitutive of agency. It is however helpful to understand this by having a discussion on what I mean when I talk about value-conferral.

Frankfurt's account of practical rationality is far too extensive and rich to do it full justice here. There are three things I take from it. First, what we care about is what guides our actions; second, what we care about provides reasons for our actions; third, caring is, somehow, a "creator" of inherent, or non-instrumental, value.

First, Frankfurt (1988) argues that reasons for action lie in what is important for ourselves and that what a person cares about directly coincides with what *guides* a person's actions. It is in virtue of us conferring value upon things that this value is somehow normatively binding over our actions. A person who cares about something does so because he or she is invested in that thing. A person

identifies himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending upon whether what he cares about is diminished or enhanced (Frankfurt, 1988, p. 260).

---

<sup>91</sup> Berys Gaut understands constructivist views as "constructivist just in case they hold that what makes something good is that it is the object of rational choice" (Gaut, 2003, p. 164). Therefore, any rational choice is value-conferring and for the constructivists who argue that, in virtue of being agents, our actions just are rational – that's what it means to be an agent – then actions themselves are value-conferring. We have already seen that rationality plays an important part in many of the constructive views discussed so far. And so, this constitutivist view has a stronger chance of helping us show why someone should care to be motivated by what they establish as valuable. Without this basic constructivist understanding of value-conferral stemming from an unconditional value, such as rational choice, the idea of value-conferral seems intangible.

Some things the agent cares more about than others, and this can be construed as one thing holding a higher value for an agent than another thing.

Second, in *The Reasons of Love* (Frankfurt, 2004), Frankfurt argues that “it is by caring about things that we infuse the world with importance” (Frankfurt, 2004, p. 23). The deepest kind of caring, according to him, is that of love and that it is in light of what one loves that our actions, as agents, are those that we want to do – thus providing reasons for actions. So, for example, I love philosophy, and because of this, I have reason to work towards being a philosopher rather than being a trapeze artist.

Third, according to Frankfurt, “insofar as love is the creator both of inherent or terminal value and of importance, then, it is the ultimate ground of practical rationality” (Frankfurt, 2004, p. 56). Because I love philosophy, I should want to be a philosopher rather than a trapeze artist. It’s also more likely that my actions and decisions will be aimed at satisfying this love I have for philosophy.

What’s interesting about Frankfurt’s theory is that we, as agents, care about objects, relationships and events in our lives and that we care insofar as we identify ourselves with what it is we care about. What’s even more interesting is that it is possible that the extent to which we care determines how valuable that object of our caring is. In other words, the more I care about being a philosopher rather than a trapeze artist, the more valuable being a philosopher is, to me. If I absolutely despise being a trapeze artist, things which are associated with trapeze artistry are not valuable to me in the least.

Back to the worth-drive then. The worth-drive is an inclination or disposition to reason or act in a certain way, such as aiming to establish that continuously choosing to be an agent is a worthwhile thing to be doing. So, conferring value onto actions etc. is something the worth-

drive should ensure happens, and happens well. To ensure this occurs without a glitch, accepting rationality norms is a constitutive requirement. So much so that without the capacity to accept rationality norms, one cannot, and should not, be considered a full-blooded agent. Third, that there is something non-instrumentally valuable about agency, and about reinforcing the continual decision to be an agent.

### 7.3.3 High-level Choice and the Worth-drive

There are a couple of things I help myself to from Frankfurt's account to help explain the worth-drive account of elective agency.

First, when I say the worth-drive aims at establishing value, I mean that we, as elective agents, are driven to establish value in those things which we care about. By choosing agency, we confer value on agency. I showed, in §5.3.1, that we can think of voluntarist reasons (in choosing agency) in terms of objective and subjective value. Voluntarist reasons are subjectively valuable insofar as they apply to me in a given situation. When I generate voluntarist reasons to choose agency, I confer value on agency. If we think of value in terms of caring, then it is that I care about agency. Caring about something means being invested in its well-being. In making sure this is something I *continue* to care about.

For example, I purchased a Samsung cell phone one year ago. It had objective value insofar as it was worth the price I paid for it. It had subjective value insofar as I liked its look. I found myself, as the months pushed on, beginning to not care as much about the phone as when I first started out. This could be because I knew I was going to renew my contract after a year and half for a new, brighter version. For whatever reason, I became increasingly more and more careless with the device until one day I cracked the screen so badly the hardware was exposed. At this point, it still worked. But I also realised that I no longer wanted this phone.

I believe that if I had a phone-worth-drive (if there was such a thing) of which the constitutive aim was to seek out ways to establish phone-value, I probably wouldn't have been careless with the device. I probably would not have broken the screen. This phone would still be my primary phone of use. But because I don't have a phone-worth-drive, I don't care about my phone much. Because of this, I don't confer the value on the phone in a way which would ensure its longevity and continued use.

If caring is something that creates inherent value, then it is important that we care about what it means to be an agent. Here, I am not talking about self-care. I'm talking about caring about the capacity to act meaningfully and voluntarily in the world. But why should this be something we should care about? Why not care about shmactions in the world instead? First, because, essentially, you can still act in the world as a non-agent, but those actions would mean absolutely nothing to you – it just seems like a pointless pursuit and I'm unsure you'd care at all. Second, you chose agency and through choosing we confer value, and we care about those things we choose – especially at the higher-level, as happens to be the case when choosing agency.

Let me return to my cell-phone example. If I had a phone-worth-drive, I would have cared about my cell phone. I would have cared if it was able to receive calls properly and if the touch screen worked as it should. Maybe I might have also cared that this is my only means to connect with my family overseas and so this device is my connection to them. This connection is no longer entirely possible with this device: the broken screen would often leave tiny shards of plastic embedded in my fingertips. So, if this cell phone is to perform the functions it was designed for, it needs to have all the parts functioning in the right way. I care that it does and so I create value – this thing is valuable to me because I care about it.

How is this important for grounding rationality norms? As I have already discussed, for an agent to identify what is valuable to her, her actions, and also her deliberations, should be governed by rationality norms. If caring is that which creates value, then it is important that there is a set of norms, in particular, rationality norms which can ensure that this value-creation is always logically and rationally guided, because, as I explained in §5.4.2, when we choose something, especially a high-level choosing, we confer value. When we confer value, we generate or create voluntarist reasons. Reasons are governed by rationality norms.

#### 7.3.4 *Recognitional Model of Value vs. Value-conferral*

Frankfurt's account establishes how value-conferral takes place. Some, however, may object to treating choice as value-conferring.

Gaut, for instance, disagrees with the idea that value is purely an object of rational choice and that rational choice is an unconditional value which holds over all other values. Gaut offers an alternative view - a recognitional model of practical reason (Gaut, 2003, p. 179). For Gaut, it is not accurate to say that action is value-conferring insofar as these actions are guided by rationality norms. Rationality norms do not, on their own, motivate us to act, and constitutive accounts fail to accommodate the notion of *human flourishing*. According to Gaut, value cannot be conferred solely under the guidance of rationality norms, instead, we also need to take into account the conception of *need* which provides the reasons for actions. According to Gaut, to figure out what it is which motivates an agent to act, we need to determine what the agent needs, but further than that, needs to what end?

Gaut argues that Kantian-type constitutivist accounts (such as proposed by Korsgaard and Velleman) are unpromising in virtue of the fact that they establish action as aimed at some kind of value or good. Instead, Gaut claims that we should be looking at recognitional models of



value. The account Gaut offers sees action as responsive to value rather than value-conferring and that value-responsiveness is not based primarily on rational choice, but also on this conception of a desire for human flourishing. Gaut offers an account that sees a *desire* or *need* for human flourishing as the reason behind action.

I argue that to think that value must be either seen on the recognitional model or on the value-conferral one is to set up a false dichotomy. In §5.3.1 I distinguished between reasons which have objective value, and reasons which have subjective value. We cannot subsume both types of reasons, I now want to argue, under the same kind of account of value. In particular, when we think about given reasons to inform our decision-making, these are the objective reasons which objectively apply to everyone. These types of reasons are better accounted for by the recognitional model of value. Voluntarist reasons, in contrast, need to be seen on the conferral model.

Under the recognitional model of value, an agent is responsive to value that exists independently of her, objective value. On the value-conferral model, we confer value on objects, decisions, actions etc. These cannot be treated in the same way. A PhD is objectively valuable insofar as it means more intellectual gratification, the advancement of a discipline, better employment, more clout in academia, and more societal respect (for example). These objective values give me objective reasons to pursue a PhD. I am responding to the objective value that is already out there.

But perhaps there are also objective reasons for me not to want to pursue a PhD. Perhaps, I encounter moments throughout the project where I begin to weigh up reasons to continue or to stop with the project. I take into account how difficult it is, what it's doing for my mental health, how it's impacted my relationship with my family for instance. I compare them to the

other given reasons to continue, and I begin to run out of reasons to do either: continue or stop. Under the recognitional model, I have given reasons to continue and given reasons to stop. And I am stuck.

The value-conferral model allows me to generate a voluntarist reason to continue. Maybe I love the sound of the clicking of my fingernails on my keyboard and know that I'd have to hear that *a lot* when writing a PhD. Maybe I generate that as my voluntarist reason to choose to continue doing the PhD. This is almost certainly not an objective reason to push forward with a PhD. But it is a subjective reason for me to do so. Not everyone shares this voluntarist reason.

It appears then that on the recognitional model, we all know the reasons why I should push through to the end, but also why we shouldn't. It's really only on the value-conferral model that I have my own reasons for wanting to push through to the end. It's possible then that given reasons could be treated as objective reasons and voluntarist reasons could be treated as subjective reasons. But presumably at least some of them could also give rise to objective, given reasons. Thus, for instance, once I have taken the nail-clicking consideration as a reason to continue my PhD, I generate a host of given, objective reasons in favour of working hard towards it, taking proper care of myself, etc. And so, we should reject the false dichotomy driving the thought that we must treat value and reasons *either* on the recognitional model *or* on the value-conferral model.

### 7.3.5 *What I Take from This*

The elective agent is one who continuously chooses to be agent. This high-level choice is where ultimate value lies. This affords us the unique position to show why it is we care to be agents – it is something we choose. Since accepting rationality norms is constitutive of the worth-drive, and the worth-drive is constitutive of agency, by choosing agency, we are choosing to

accept norms rather than shnorms. This puts us in the unique position to establish the normative grip of rationality norms.

So far, I defined the worth-drive as a functional state which motivates an agent to act in a way that validates the choice of being an agent. I then discussed the way Frankfurt uses value in his account. I would like to use the following ideas from Frankfurt's position to establish how the worth-drive works: a) action as value-conferring; b) what we care about creates reasons which normatively and motivationally guide our actions.

Action, understood as value-conferring is an enterprise unique to full-blooded agency. Valuing and value-conferring is not a random exercise performed by chance, but instead is closely dependent on the acceptance of rationality norms. This is what distinguishes it from mere plumping for one option rather than another. There is no value-conferral for the individual who wishes to determine the best course of action by flipping a coin – this has been discussed extensively in the last chapter. Agents, uniquely, hold a value-conferral status in virtue of their capacity to accept rationality norms.

#### **7.4 The Final Shtest**

For a constitutive account of agency to successfully sustain the normative grip of rationality norms, it needs to pass the Shtest. First, it needs to be able to explain purposeful norm-violation without compromising agency. Second, it should not appeal to the unavailability of agency to answer the why-care question. Third, what is identified as the constitutive feature of agency must be shown to be non-arbitrary and to enjoy normative-uniqueness. Fourth, the account of agency should be able to answer questions about agency without appealing to self-vindication.

##### *7.4.1 Accommodating Norm-violation*

Most constitutive accounts show how norm-violation is only possible in the weak sense. The reason for this is their commitment to the idea that to be an agent is to be a thing which adheres to norms, but further, that to be an agent is unavoidable. I argue, that to be an agent is to be a thing which accepts norms, but that this acceptance of norms is chosen. By introducing genuine choice, insofar as we even choose our agency, we are in the position to explain norm-violation, and thus do not rule out genuine norm-adherence. This I have argued for in the previous two chapters. I have also shown how it is, that because of the worth-drive, we act in a way that confirms that the choice we make to be agents is a worthwhile one.

Normativity as an unavoidable constitutive notion (as put forth by Korsgaard and Velleman), is something we have no choice over; it is something that is a part of our constitutive make-up and thus lacks the force we require to ground the source of normativity. In contrast, agency which is embedded in choice – as is for the elective agent - allows for normativity. Think again of the discussion about norm-violation. True norm-adherence requires the capacity or ability for norm-violation. If we don't have a choice about X, then X cannot be a source for what we ought to do, because what we ought to do presupposes that one is able to follow or violate X.<sup>92</sup> Accounts of agency that dictate that we cannot but be agents are stuck with norm-adherence in the false sense. The elective agent, however, is not stuck with this type of restriction or constraint and so, the capacity for norm-violation necessary for norm-adherence is in place.

How does the worth-drive allow for norm-violation? First, because accepting rationality norms is constitutive of the worth-drive, when one chooses to embrace the worth-drive, one chooses to accept rationality norms. But, because this is a choice to accept rather to always adhere to, we are in the position to either adhere to the norm or to violate it. Furthermore, because

---

<sup>92</sup> I argued for this in chapter 2.

acceptance is involved, the violation in question is typically strong violation, the kind I argued a constitutive account must be able to accommodate.

Two things to note here. Adhering to norms will have better outcomes when it comes to validating our choice to be an agent. This is an instrumental connection. But there is also a non-instrumental one: only elective agents are those things which embrace the worth-drive and only elective agents are those that seek out actions which validate the choice to be an agent. Violating norms enough of the time has the capacity to diminish the value of being an agent and to diminish reasons to continue to choose agency. Because we are inclined to perform those actions which validate our choosing to be an agent.

#### 7.4.2 *The Unavoidability of Agency*

The worth-drive overcomes the problem of unavoidability in obvious ways. Agency here is *not* something we just find ourselves doing and should thus care about. On the contrary, the *choice* to be an elective agent is one of the distinguishing marks of the proposal. An individual chooses to be an agent when they embrace the worth-drive. The worth-drive is that functional motivational drive which pushes an agent to seek out worth in a way that validates her decision to be an agent.

In this way, if an agent does enough ugly actions which diminishes the worth of being an agent, she may choose to opt out of agency. If she performs actions which establish the worth of being an agent, she will continuously opt into agency. Thus, the problem of unavoidability is not a concern for the worth-drive.

#### 7.4.3 *Non-arbitrariness*

Remember Enoch's concern? "The question was why it is that being constitutive of agency renders these motives and capacities [that are supposed to ground normativity] non-arbitrary" (Enoch, 2006, p. 43). The answer the account of elective agency gives is that the non-arbitrariness does not hinge on these motives and capacities being constitutive of agency. Non-arbitrariness under this account is established in virtue of us choosing this particular drive. Only agents are those things which embrace the worth-drive, which means the worth-drive is unique to elective agency. Second, by embracing the worth-drive, we accept rationality norms, which ensures normative-uniqueness of this constitutive feature of agency.

What allows the worth-drive to hold its privileged normative-status is that, under the guidance of rationality norms, it ensures actions are aimed at establishing value in being an agent. This drive is non-arbitrary as it is exclusive to agency and serves only agency.

#### *7.4.4 Self-vindication*

In the previous chapter, I argued that there are elements of elective agency which are similar to other enterprises, such as the enterprise of playing chess. The similarity lies in our ability to choose to play chess as well as to engage in this enterprise of elective agency. Second, that just as we choose to accept the norms associated with chess-playing and we further accept that the norms of chess-playing will guide us in how to play the game to the best of our abilities, we choose to accept norms associated with agency, through embracing the worth-drive. We also accept that these norms will guide us in a way that we would want to continue being agents in virtue of adhering to these norms.

I have already shown here that the worth-drive is a particular drive which only agents have, and this is a kind of drive distinctly different from other types of drives. Under the account of elective agency, there are two possible ways to answer the self-vindication constraint:

i) The worth-drive is something that drives us to continue choosing agency over non-agency.

This enables us to evaluate the worth of being an agent from a place of agency which is essentially of our choosing, evaluating it and asking of it if it is worth it to continue in this enterprise.

ii) perhaps it is possible to show that even though elective agency itself falls into the trap of self-vindication – and even though this is no longer a problem if severed from the problem of the unavoidability of agency – that once we flesh out this account in terms of the worth-drive, we actually are able to move beyond the problem of self-vindication (independent of severing the link). How is this possible?

Let's go back to the Geraldine example. When Geraldine is unable to lift the tree from the elderly lady because she is terrified of the termites – she is *not* being irrational. I think it's quite rational to avoid doing something that revolts you. But what Geraldine is doing here is diminishing the value of being an agent – essentially making a worthless choice. It's worthless in the sense that she has considered an important reason (to save a life), and she has considered a less important reason (her revulsion of termites) and has added more weight to the less important reason than what she should have.

Here, she is adhering to rationality norms. But if asked, Geraldine might say that this a norm she doesn't readily accept. When it comes to norm-acceptance, she would have rather liked to save the life than to be revolted by termites. She accepts that life-saving is a great norm to follow, but she just doesn't adhere to this norm in this instance. Instead, she adheres to a norm that she doesn't accept – still rational however. Here, she is still adhering to norms, but she is frustrating the constitutive aim of the worth-drive. In the same way, perhaps it is possible that

when I am asked why it is that I should care about following rationality norms, I don't need to employ the worth-drive to answer this question.

## 7.5 Objections

Let me consider four possible objections by way of strengthening my proposal. The first objection deals with the idea that norm-violation cannot actually take place under the worth-drive, as advertised. The second possible objection is that perhaps there is something else other than the worth-drive that can do the same job. Third, whether the worth-drive is descriptive or normative. Finally, whether the worth-drive isn't the most parsimonious story to be telling about elective agency.

### 7.5.1 *No Actual Normative Grip*

The first possible objection is that the worth-drive does not allow for norm-violation as advertised. If we have accepted rationality norms and the worth-drive aims actions at establishing value in being an agent, then it's unclear how we are able to violate norms.

My response is that I have already argued (in §7.4) that actions can either undermine or confer value. Both instances are equally possible. However, if actions undermine the choice to be an agent often enough, an agent may be driven to choose something else instead, such as non-agency, perhaps shmagency (if that were an option). By the action of choosing, we are endorsing rationality norms. But, endorsing something does not mean automatic adherence. Here, I provide an example.

At times, it is possible that the worth-drive will identify value in holding two contradictory beliefs and this happens because an agent is not able to process the truth of one belief and



chooses another instead. This is an instance of genuine norm-violation. For instance, hearing the terrible news that a loved-one has stage 4 cancer could potentially render an agent with two contradictory beliefs: i) the loved-one is most likely going to die and ii) the loved-one will most likely survive. Worse still, even in the face of evidence – such as the news that the cancer has spread to the brain – may still leave the agent holding the belief that everything is going to be fine. This kind of norm-violation affects action. Perhaps, the agent, hearing the news that the loved-one won't live longer than three months, plans a birthday party for the love-one 10 months ahead. This kind of behaviour is not uncommon. The worth-drive establishes a value for the agent in holding this contradictory belief, such that if the agent did not hold this contradictory belief, the agent may reconsider embracing agency.

#### *7.5.2 Anything Else other than the Worth-Drive?*

A second possible objection is that perhaps there is something else an individual can embrace to become an agent. Such as Velleman's drive for sense-making. Why the worth-drive specifically?

My response is that I have already shown in Chapter 4 that Velleman's account fails the Shtest for several different reasons, precisely because the drive for sense-making assumes the unavailability of agency to begin with. But, let's imagine for a moment that Velleman did not commit to the notion of the unavailability of agency, and that he too commits to the idea that agency is something one can choose, would the drive for sense making achieve what the worth-drive does?

The answer is, as far as I can tell and with a bit of imagination, maybe. Firstly, with the worth-drive, things are very different. Because higher-level choice is what confers value, choosing agency confers ultimate value. The worth-drive is that which continuously pushes an agent to

pursue actions which confer value in a way that validates the choice to an agent. If you perform actions enough times that diminish the value of being an agent, you are driven to question your choice to continue choosing agency over non-agency.

It just seems, at least intuitively so, that this is just not something another drive can give us. What would need to be done in order for the drive for sense-making to work in the same way as the worth-drive, is to establish that to choose agency is to embrace sense-making or something like it. That somehow, when we choose agency, the choice itself is an instance of ultimate sense-making, such that all other actions aim at confirming that this choice is what makes the most sense to do and that all other actions fall under this umbrella choice. But, for something to make sense, it must still appeal to value of some sort. And here, I'm worried that the drive for sense making is unable to confer value. I maintain that the worth-drive is the best candidate to accomplish this.

### 7.5.3 *Is the Worth-drive Normative or Descriptive?*<sup>93</sup>

Some may worry at this point that I face a dilemma. On the one horn, the connection between the worth-drive could be seen as purely descriptive. But then how could this connection ground the normativity of anything? On the other horn, if it is normative, then it falls prey to Enoch's objection and the problems set out in the Shtest. If the worth-drive is as I characterise it to be, then it appears to be a drive which all agents have (descriptive). But it is also something we supposedly have reason to embrace (normative). Both are problematic.

---

<sup>93</sup> Many thanks to Christoph Hanisch who made me aware of this objection.

I answer this objection in the following way: we can understand the worth-drive as having both a descriptive and a normative dimension, it is to use Williams's notion, a 'thick concept'. I now show how this gets me around this dilemma.

In *Ethics and the Limits of Philosophy* (1985), Williams introduced the term 'thick concepts' where certain terms or concepts are both evaluative and descriptive, in a way that the two dimensions cannot be disentangled from each other. For example, consider the term 'rude'. According to Philippa Foot, the term 'rude' "expresses disapproval, is meant to be used when an action is to be discouraged, implies that other things being equal the behaviour to which it is applied will be avoided by the speaker, and so on" (Foot, 1958: 102). Descriptively, the term 'rude' also describes an action or behavior which is essentially an action one *should* avoid in social settings.

I suggest that the worth-drive is similarly, both descriptive and normative. The worth-drive, as a concept, refers to a drive which strives to establish value in the world through certain actions, therefore, because one has the worth-drive, one ought to behave the way the worth-drive prescribes. Every agent has the worth-drive and its job is to guide our behaviour such that we are always striving to adhere to rationality norms. But Enoch may, at this point, say that if the drive is some kind of inescapable drive or inclination, then what is the normative status of the worth drive?

Note, that this drive is not inescapable. It is a drive you embrace when you choose agency and so does not fall into the trap of unavoidability. But why care to choose it? As a thick notion, the worth-drive is both descriptive, it is something an agent has, and it is normative, it is something that one should embrace. And so, when we choose agency, we choose that which is tied to agency – such as the worth drive. Instrumentally, the worth-drive will serve you well as

an agent in the world, as it will ensure you strive towards establishing worth as far as you go. The worth-drive is a vehicle for rationality-norms. It is the gravity that ensures the packet of diamonds remains safe (see chapter 2 for this example as pertaining to constitutivism). Non-instrumentally, the worth-drive can be understood as the source of subjective (and objective) value. Instrumentally and non-instrumentally, the worth-drive is worth embracing because the worth-drive underlies an agent's value-conferring activities, it serves as both a motivational force (descriptive) and guides our actions (normative).

#### *7.5.4 The Parsimony Objection<sup>94</sup>*

Surely it's not necessary, the reader might say, to have this story of the worth-drive. We can imagine a simpler agent who can make choices, without having to appeal to the worth-drive to tell this story. It's surely enough to just have rationality-norms guide our behaviour rather than positing an additional component: the worth-drive.

My response to this is that perhaps there is a simpler story to tell and perhaps the worth-drive is not the only story that can be told about how it is that we are guided by rationality norms. Indeed, I provided such a simpler story in chapter 5 and 6 in the account of elective agency. I had only introduced a black box notion without providing a possible mechanism as to how elective agency would work. The worth-drive is one way to flesh out that story and provide a possible mechanism as to how elective agency is possible. This is a deeper explanation as to why, through high-level choice, we choose and continue to choose agency; it also provides an account of what our actions aim at. Notice though that my answer to the why-care question does not depend on this story of the worth-drive. It's enough to say that we aim our actions at establishing worth, for without the worth-drive, our agent may be rudderless. But perhaps a

---

<sup>94</sup> Many thanks to Jessica Lerm for making me aware of this objection.

challenge here would be to say that the agent would not be rudderless without the worth-drive, because they'd have rationality norms to guide them. But where is the story of motivation when we have rationality norms hovering over us and no account as to why we would feel motivated to be guided by them. The worth-drive is one potential answer to this.

## **7.6 Conclusion**

In this chapter, I provided a substantive account of what elective agency is by introducing a drive constitutive of agency: the worth-drive. I first unpacked what I mean by a drive, establishing it as something which consistently motivates us to aim our actions and deliberations at value. I establish that a key notion to elective agency is that accepting rationality norms is constitutive of the worth-drive. I show how the worth-drive would not function optimally if not under the governance or guidance of rationality norms. I then showed how agents are those things which embrace the worth-drive and further showed that we are elective agents in virtue of embracing this worth-drive.

I proceeded to unpack the idea that the worth-drive aims at identifying and establishing value in actions and in agency itself. To illustrate what I mean when I speak of value, I discussed the distinction between instrumental and non-instrumental value.

I then discussed Frankfurt's account of value which establishes that we value things we care about and that actions are value-conferring. Frankfurt further argued that love is a creator of value, arguing that value-conferral and value-responsiveness are closely linked to the acceptance of rationality norms, thus tying in the idea that accepting rationality norms is constitutive of the worth-drive. I then proceeded to show that the elective agent, in virtue of the worth-drive, fares far better than any of the other accounts with regards to the Shtest.



UNIVERSITY  
OF  
JOHANNESBURG

## Chapter 8 – Conclusion

### 8.1 Synopsis

Accounts that wish to ground the source of the normativity of rationality encounter various challenges, most notably, why it is that we should adhere to norms rather than violate them. When it comes to rationality norms, it just seems reasonable that adhering to rationality norms then would be a beneficial thing to be doing. But note, that we can't run that argument, firstly because of its circular nature. Second, because why should an agent care that something is a beneficial thing to be doing? Even if we could establish why something being beneficial means it is worth doing, it is further difficult to establish that adhering to rationality norms will be what gets you there.

A potential fix is to situate the normative grip of rationality norms in a constitutive account of agency – resulting in the overall conclusion that the normative grip of rationality norms can be found in what is constitutive of agency, for example, beliefs, desires or drives. This is a potential fix because, if accepting rationality norms is constitutive of this feature of agency, then, in virtue of being an agent, we are guided by rationality norms. Problematically though, even though current constitutive accounts of agency adopt this strategy, they do not actually achieve the aim of grounding the source of rationality norms. This failure can be attributed to two core issues. First, all the accounts explored in this project too intimately link the problem of the unavoidability to agency and the problem of self-vindication. Second, none of the accounts are able to successfully answer the why-care question. First, why we should care to follow rationality norms rather than anything else, and second, why we should care to even be agents as opposed to anything else.

The aim of this project was to determine what it would take to successfully ground rationality norms in a constitutive account of agency. In order to do this, I first discussed certain problems that need to be overcome in order for a constitutive account of agency to be in the position to show how rationality norms have their normative grip over our decision-making and actions. Three out of the four problems are challenges Enoch puts forward to constitutivists coupled with a challenge I put forth of my own. I call this the *Shtest*. I examined four different constitutive accounts of agency and concluded that all four fail to successfully ground the source of rationality norms. I proceeded then to identify that the core issue as to why this failure occurs in every account is because each account assumes that agency is unavoidable. I proceed to provide a new constitutive account of agency I called *elective agency*. The account proposes that agents actually have full choice even when it comes to matters about choosing to be agents over something else, like non-agency, or in some instances, shmagency. I argue that the way an elective agent is able to choose to be an agent is through embracing the worth-drive.

### *Chapter 1: Introduction*

Chapter one served to provide a brief and general contextualisation of theories of action, agency, normativity and caveats. Finally showing the crux of the problem regarding sustaining the normative grip of rationality norms. How is it that we are compelled to be governed by rationality norms? To show how this is possible and how rationality does essentially have its normative grip, constitutivists have attempted to ground the source of rationality in what is constitutive of agency. Essentially, grounding rationality norms in the feature which is constitutive of agency ensures that you adhere to rationality norms in virtue of being an agent.

What is problematic about this is that things begin to unravel when you ask the question of why we should care to adhere to these norms rather than, say, shnorms (throughout the project,



I referred to this as the why-care question). To answer this, constitutivists usually appeal to the notion that we care in virtue of being agents – that’s just what being constitutive means. And that answer seems to stand its ground until you ask the constitutivist why we should care to be agents. I indicated that answering this question will essentially be the focus of this project.

### *Chapter 2: The Shtest*

In Chapter two, I set out the challenges encountered by those wishing to ground the source of rationality norms. Here, I explained different approaches to normativity as well as different camps that attempt to establish the normative grip. It turns out however that it doesn’t matter which camp you find yourself supporting, constitutive accounts fail the Shtest.

The failure looks something like this: constitutive accounts of agency insist that the feature they identify as constitutive of agency somehow holds a privileged normative-status – in virtue of being constitutive of agency – which essentially is supposed to answer why agents feel compelled to be rational given the normative grip this feature apparently has. But according to Enoch, just because something is a constitutive feature, does not mean it automatically holds a privileged normative-status. Because, as Enoch points out, why is agency so compelling in the first place?

I briefly outlined the four problems here. First, the *problem of norm-violation*. Second, the *problem of the unavailability of agency*. Third, the *problem of non-arbitrariness*. Finally, the problem of *self-vindication*.

The purpose of the Shtest is to determine if constitutive accounts are able to sustain the normative grip of rationality norms. I proceeded, in chapter 3 to shtest two constitutive accounts of agency.

### *Chapter 3: Constitutive Features: Actions or Beliefs?*

In chapter three, I examine two prominent constitutive accounts of agency. The purpose of this chapter is to see how well the two accounts perform with regards to the Shtest.

Firstly, I examine Korsgaard's account which pushes for the notion of self-constitution. I show that Korsgaard's account does not fare well in the Shtest. First, Korsgaard account is unable to show how norm-violation takes place without either diminishing agency or terminating it entirely. Second, Korsgaard is the strongest advocate that agency is unavoidable and that it is something we are essentially condemned to. Third, Korsgaard's account does not adequately show what it is that makes a particular feature of agency non-arbitrary. Finally, Korsgaard wholeheartedly commits to the idea of self-vindication, because to ask the why-care question is to act, which according to Korsgaard is self-constituting.

Railton does a bit better comparatively. He shows how norm-violation takes place, at least in the weak sense by stating that a belief can be wrong if it misses the mark it was essentially aiming for. Second, Railton also, unfortunately commits to the unavoidability of agency. Third, Railton shows, to a degree, how the constitutive feature is non-arbitrary but also, unfortunately commits to the problem of self-vindication by showing that in the process of asking the question, one is appealing to that which is constitutive of agency to do so.

### *Chapter 4: Constitutive Features: Drives?*

In chapter four, I discussed two further constitutive accounts of agency put forward by Velleman and by Katsafanas. However, for the purpose of this project, my primary focus was Velleman's.

Both accounts do better than the accounts offered by Korsgaard, and Railton. But, unfortunately, neither succeeds with regards to the Shtest. In both instances, norm-violation is possible at the cost of frustrating the drive. But both accounts commit to the unavoidability of agency. Both accounts run up against the problem of self-vindication.

I concluded this chapter by providing a diagnosis: the critical flaw of constitutivist accounts is to link too intimately the notion of unavoidability and self-vindication. Insofar as they do so, they will always fail the Shtest. The reason for this is that linking them in this way blocks you from explaining what is non-arbitrary and unique about your chosen constitutive feature. This is a problem my constitutive account of agency avoids.

#### *Chapter 5: Elective Agency*

In this chapter I put forth my positive contribution to the debate - Elective Agency. The account of elective agency fully embraces the idea of choice in the genuine sense. Since choice is considered a hallmark feature of agency, it just seems right to incorporate the notion of choice into all aspects of agency, including choosing agency over non-agency or shmagency. I argued that an elective agent is an agent who is able to choose, and continuously choose to be an agent.

In this chapter I developed a black-box notion of elective agency without providing the nitty-gritty mechanisms foundational to this account – that I left for chapter seven.

#### *Chapter 6: Elective Agency and the Shtest*

I argued that this black-box notion of agency does better to answer the problems of norm-violation, the problem of the unavoidability of agency, the problem of non-arbitrariness and the problem of self-vindication.

I argued that norm-violation is possible under this account because an elective agent can choose agency, and thereby can choose the norms associated with it. Second, if agency is something we can choose, then it can be avoided – thus not falling up against the unavoidability problem. With regards to non-arbitrariness, because accepting rationality norms is constitutive of agency then by choosing agency, we are choosing to adhere to rationality norms, thus establishing non-arbitrariness. Finally, with regards to the problem of self-vindication, because elective agency does not intimately link self-vindication with the unavoidability of agency, this proves to not be a problem.

### *Chapter 7: The Worth-Drive*

This is the third positive chapter in this project, where I unpack how elective agency is possible. First, I established that what is constitutive of agency is something I call the worth-drive. I identified this drive as that which motivates us to aim our actions and decision-making at identifying and conferring value. The fundamental argument here is that we, as elective agents, embrace the worth-drive and is this that sets us apart from shmagents and non-agents. When an individual embraces the worth-drive, an individual accepts rationality norms

The worth-drive aims at value. But, I argued that even if action is value-conferring, it can also be value-diminishing. This is central to the notion of the worth-drive. Because action can be both, the role of the worth-drive is to motivate us to choose actions which are value-conferring, not value-diminishing. To do this, actions need to validate our decision or choice to be an agent since agency is something we essentially embrace. If we approve of the actions we perform in light that this is the type of agent I confer value on, then I continue choosing agency over shmagency or non-agency. I concluded that the constitutive account of elective agency, in virtue of embracing the worth-drive, is better able to meet the problems set out in the Shtest.

## 8.2 The Contribution

The unique contribution this project aims to make to debates on grounding the source of rationality norms is a new account of agency which can successfully answer two core concerns, and thus ground rationality norms. This new account of agency centres on two core features. First, that we have *chosen* and continue to *choose* to be agents – I call this *elective agency*. Through choosing to be agents, we are in a unique position to answer questions related to *why we should care* to be agents rather than non-agents. Second, as elective agents, we are in possession of the *worth-drive*. By embracing the worth-drive, we seek to establish the value of agency, of being an agent, and of choosing agency. It is through embracing the worth-drive that we elect agency.

I further argued that accepting rationality norms is constitutive of the worth-drive and claimed that the worth-drive is constitutive of agency. Because we choose agency, we choose to accept the rationality norms that come along with it. We accept these rationality norms in a way that can allow for norm-violation – which other accounts fail to do. I argue that this constitutive account puts us in a unique position to ground the source of the normativity of rationality norms.

## 8.3 Further Study

A potential future study resulting from this project could be to explore the relationship between the worth-drive and normativity in a more general way. This project focused primarily on rationality norms and sustaining the normative grip of rationality norms. What would be interesting to see is if norms pertaining to beauty, for instance, can be grounded in such a constitutive account of agency. Furthermore, if a constitutive account, such as outlined in elective agency, can somehow explain norms regarding wisdom, meaning in life, mental health,

treatment of the environment and animals etc.<sup>95</sup> Perhaps, and rather ambitiously so, a constitutive account of agency of this kind could be able to provide some sort of unified theory of normativity. But this ambitious project I leave for another day.



---

<sup>95</sup> Thank you to Thaddeus Metz for mentioning this in a review.

## References

Anderson, E., 1993. *Value in Ethics and Economics*. Cambridge : Harvard University Press.

Anscombe, G.E.M. 1957. *Intention*. USA: Harvard University Press.

Audi, R., 2001. *The Architecture of Reason: The Structure and Substance of Rationality*.

Oxford: Oxford University Press.

Audi, R., 2003. Moral Judgement and Reasons for Action. In: *Ethics and Practical Reason*.

New York: Oxford University Press, pp. 125 - 160.

Audi, R., 2004. Reasons, Practical Reason, and Practical Reasoning. *Ratio*, XVII(2), pp. 119-149.

Bagnoli, C., 2002. Moral Constructivism: A Phenomenological Argument. *Topoi*, 21(1), pp. 125 - 138.

Bagnoli, C., 2017. *Constructivism in MetaEthics*. [Online]

Available at: <https://plato.stanford.edu/cgi->

[bin/encyclopedia/archinfo.cgi?entry=constructivism-metaethics](https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=constructivism-metaethics)

[Accessed 15 October 2018].

Balaguer, M., 2009. Why There are no Good Arguments for any Interesting Version of Determinism. *Synthese*, Volume 168, pp. 1 - 21.

Balaguer, M., 2014. Replies to McKenna, Pereboom, and Kane. *Philosophical Studies*, Volume 169, pp. 71 - 92.

- Bishop, J., 1989. *Natural Agency: An Essay on the Causal Theory of Action*. New York: Cambridge University Press.
- Boghossian, P., 2003. Blind Reasoning. *Proceedings of the Aristotelian Society Supplementary Volume*, 77(1), pp. 225-248.
- Brand, M., 1979. The Fundamental Question in Action Theory. *Nous*, Volume 13, pp. 131 - 151.
- Brand, M., 1984. *Intending and Acting*. Cambridge: MIT Press.
- Bratman, M., 2001. Two Problems about Human Agency. *Proceedings of the Aristotelian Society*, Volume 101, pp. 309 -326.
- Bratman, M., 2006. Shared Valuing and Frameworks for Practical Reasoning. In: R. Wallace, P. Pettit, S. Scheffler & M. Smith, eds. *Reason and Value*. New York: Oxford University Press, pp. 1-27.
- Bratman, M., 2007. *Structures of Agency*. New York: Oxford University Press.
- Brentano, F., 1969. *The Origin of Our Knowledge of Right and Wrong*. London: Routledge & Kegan Paul.
- Broome, J., 1999. Normative Requirements. *Ratio*, Volume 12, pp. 398 - 419.
- Broome, J., 2003. Practical Reasoning. In: J. Bermudez & A. Millar, eds. *Reason and Nature*. Oxford: Oxford University Press, pp. 85 - 112.
- Broome, J., 2005. Does Rationality Give Us Reasons?. *Philosophical Issues*, 15(Normativity), pp. 321 - 337.
- Broome, J., 2008. Is Rationality Normative?. *Disputatio*, II(24), pp. 153 -170.



Brunero, J. & Kolodny, N., 2013. *Instrumental Rationality*. [Online]

Available at: <https://plato.stanford.edu/entries/rationality-instrumental/>

[Accessed 23 05 2018].

Chang, R., 2009. Voluntarism and the Sources of Normativity. In: D. Sobel & S. Wall, eds. *Reasons for Action*. UK: Cambridge University Press, pp. 243 - 271.

Cohen, G., 1996. Reason, Humanity, and the Moral Law. In: *The Source of Normativity*. Cambridge: Cambridge University Press, pp. 167 - 188.

Coleman, J., 1990. *Foundations of Social Theory*. Cambridge: Belknap Press of Harvard Press.

Cote-Bouchard, C., 2016. Can the Aim of Belief Ground Epistemic Normativity?. *Philosophical Studies*, 173(12), pp. 3181 - 3198.

Cullity, G. & Gaut, B., 2003. Introduction. In: G. Cullity & B. Gaut, eds. *Ethics and Practical Reason*. New York: Oxford University Press, pp. 1 - 27.

Dancy, J., 1993. *Moral Reasons*. Oxford: Blackwell.

Dancy, J., 2000. *Practical Reality*. Oxford : Oxford University Press.

Davidson, D., 1963. Actions, Reasons, and Causes. *The Journal of Philosophy*, 60(23), pp. 685-700.

Davidson, D., 1982. Paradoxes of Irrationality. In: J. Hopkins, ed. *Philosophical Essays on Freud*. Cambridge: Cambridge University Press, pp. 289 - 305.

Davidson, D., 2001. *Essay on Actions and Events*. Oxford: Clarendon Press.

Dretske, F., 1970. Epistemic Operators. *The Journal of Philosophy*, Volume 67, pp. 1007 - 1023.

Dretske, F., 1981. *Knowledge and the Flow of Information*. Cambridge: MIT Press.

Dunn, J., 2015. Reliability for Degrees of Belief. *Philosophical Studies*, 172(7), pp. 1929 - 1952.

Elster, J., 1991. Rationality and Social Norms. *Archives Européennes de Sociologie*, 32(1), pp. 109 - 121.

Enoch, D., 2006. Agency, Shmagency: why normativity won't come from what is constitutive of action. *Philosophical Review*, 115(2), pp. 31 - 60.

Enoch, D. 2011. Shmagency Revisited. *New Waves in Metaethics*. London: Palgrave Macmillan/

Feldman, R., 1985. Reliability and Justification. *The Monist*, 68(2), pp. 159 - 174.

Feldman, R., 2003. *Epistemology*. New Jersey: Prentice-Hall.

Ferrero, L., 2009. Constitutivism and the Inescapability of Agency. *Oxford Studies in Metaethics*, Volume IV, pp. 303-333.

Finlay, S. & Schroeder, M., 2017. *Reasons for Action: Internal vs. External*. [Online]

Available at: <https://plato.stanford.edu/archives/fall2017/entries/reasons-internal-external/>

[Accessed 21 12 2018].

Foot, P., 1958. Moral Arguments. *Mind*. 67 (268). Pp. 502 - 513.

- Foot, P., 1967. The Problem of Abortion and the Doctrine of the Double Effect. *Oxford Review*, Volume 5.
- Foot, P., 1985. Utilitarianism and the Virtues. *Mind*, 94(2), pp. 196 - 209.
- Frankfurt, H., 1969. Alternate Possibilities and Moral Responsibility. *The Journal of Philosophy*, 66(23), pp. 829 -839.
- Frankfurt, H., 1971. Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), pp. 5-20.
- Frankfurt, H., 1988. *The Importance of What we Care About*. Cambridge: Cambridge University Press.
- Frankfurt, H., 1999. *Volition, Necessity, and Love*. Cambridge: Cambridge University Press.
- Frankfurt, H., 2004. *The Reasons of Love*. New Jersey: Princeton University Press.
- Gaut, B., 2003. The Structure of Practical Reason. In: G. Cullity & B. Gaut, eds. *Ethics and Practical Reason*. New York: Oxford University Press, pp. 161-188.
- Ginet, C., 1990. *On Action*. Cambridge: Cambridge University Press.
- Goldman, A., 1986. *Epistemology and Cognition*. Cambridge: Harvard University Press.
- Goldman, A., 2008. Immediate Justification and Process Reliabilism. In: *Epistemology: New Essays*. New York: Oxford University Press, pp. 63 - 82.
- Goldman, A. & Beddor, B., 2016. *Reliabilist Epistemology*. [Online]  
Available at: <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=reliabilism>  
[Accessed 21 07 2019].

Haggard, P., 2005. Conscious Intention and Motor Cognition. *Trends in Cognitive Sciences*, 9(6). Pp. 290–295.

Hansson, S., 1998. *Structures of Value - An Investigation of the Statics and Dynamics of Values and Norms*, s.l.: Lund University.

Harman, G., 1975. Moral Relativism Defended. *Philosophical Review*, Volume 84, pp. 3 - 22

Henkel, L. & Mather, M., 2007. Memory Attributions for choices: How Beliefs shape our Memories. *Journal of Memory and Language*, Volume 57, pp. 163 - 176.

Hlobil, U., 2014. Against Boghossian, Wright and Broome on inference. *Philosophical Studies: An international Journal of Philosophy in the Analytic Tradition*, 167(2), pp. 419-429.

Hobbes, T., 1651. *The Leviathan*. Oxford: Clarendon 1909 reprint.

Hoffe, O., 1993. Review of Constructions of Reasons by Onara O'Neill. *Philosophische Rundschau*, 40(1-2), pp. 83 - 86.

Hume, D. 1939. *A Treatise of Human Nature*. London: John Noon.

Hursthouse, R., 1991. Arational Actions. *The Journal of Philosophy*, 88(2), pp. 57 - 68.

Jackson, F., 1982. Epiphenomenal Qualia. *Philosophical Quarterly*, 32(127), pp. 127 - 136.

Jacobson, D., 2011. Fitting Attitude Theories of Value. *Stanford Encyclopedia of Philosophy*, Issue The Stanford Encyclopedia of Philosophy (Spring 2011 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/spr2011/entries/fitting-attitude-theories/>>.

Johnson, A. & Hathcock, D., n.d. Study Abroad and Moral Development. *ejournal of Public Affairs*, 3(3), pp. 52 - 70.

Kagan, S., 1998. Rethinking Intrinsic Value. *The Journal of Ethics*, Volume 2, pp. 277-297.

Kant, I., 1785. Groundwork for the Metaphysics of Morals. In: A. Wood, ed. *Groundwork for the Metaphysics of Morals*. New York: Yale University, pp. 1 - 63.

Kant, I., 1788. The Critique of Practical Reason. In: W. Pluhar, ed. *The Critique of Practical Reason*. Indianapolis: Hackett Publishing Company.

Kant, I., 1793. Religion within the Boundaries of Mere Reason. In: A. Wood & G. di Giovanni, eds. *Religion within the Boundaries of Mere Reason*. Cambridge: Cambridge University Press.

Kant, I., 1797. The Metaphysics of Morals. In: M. Gregor, ed. *The Metaphysics of Morals*. Cambridge: Cambridge University Press.

Kant, I., 1798. Anthropology from a Pragmatic Point of View. In: R. Louden, ed. *Anthropology from a Pragmatic Point of View*. Cambridge: Cambridge University Press.

Katsafanas, P., 2013. *Agency and the Foundation of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.

Kolodny, N., 2005. Why be Rational?. *Mind*, 114(455), pp. 509 - 563.

Korsgaard, C., 1983. Two Distinctions in Goodness. *Philosophical Review*, Volume 92, pp. 169 - 195.

Korsgaard, C., 1996. *The Sources of Normativity*. O'Neill, O ed. Cambridge: Cambridge University Press.

Korsgaard, C., 1997. The Normativity of Instrumental Reason. In: G. Cullity & B. Gaut, eds. *Ethics and Practical Reason*. Oxford: Clarendon, pp. 215 - 254.

Korsgaard, C., 2000. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

Korsgaard, C., 2009. *Self-Constitution*. Oxford: Oxford University Press.

Latour, B., 2000. Opening Pandora's Black Box. In: I. McLoughlin, D. Preece & P. Dawson, eds. *Technology, Organizations and Innovation: Theories, concepts and paradigms*. London: Routledge .

Lavin, D., 2004. Practical Reason and the Possibility of Error. *Ethics*, 114(3), pp. 424-457.

LeBar, M., 2008. Aristotelian Constructivism. *Social Philosophy and Policy*, 25(1), pp. 182-213.

LeBar, M., 2013. Constructivism and Particularism. In: C. Bagnoli, ed. *Constructivism in Ethics*. Cambridge: Cambridge University Press, pp. 183 - 200.

Lenman, J., 2010. Humean Constructivism in Moral Theory. *Oxford Studies in Metaethics*, Volume 5, pp. 175 - 193.

Lenman, J., 2012. Expressivism and Constructivism. *Constructivism in Practical Philosophy*, Issue doi:10.1093/acprof:oso/9780199609833.003.0012, pp. 213 - 225.

Lindeman, K., 2017. Constitutivism without Normative Thresholds. *Journal of Ethics and Social Philosophy*, 12(3), pp. 221 - 258.

Littlejohn, C., 2012. *Justification and the truth-connection*. New York: Cambridge University Press.

Locke, J., 1689. *An Essay Concerning Human Understanding*. In: P. Nidditch, ed. *An Essay Concerning Human Understanding*. Oxford: Clarendon Press.

MacDowell, J., 1994. *Mind and World*. Cambridge: Harvard University Press.

Mackie, J., 1977. *Ethics: Inventing Right and Wrong*. s.l.:Harmondsworth: Penguin Books.

Magill, K., 1997. *Freedom and Experience: Self-determination without Illusions*. USA: St. Martin's Press.

McDowell, J., 1978. Are Moral Requirements Hypothetical Imperatives?. *Proceedings of the Aristotelian Society*, Volume 52, pp. 13 - 29.

McHugh, C., 2013. Epistemic Responsibility and Doxastic Agency. *Philosophical Issues*, 23(1), pp. 132-157.

Mele, A. & Moser, P., 1994. Intentional Action. *Nous*, 28(1), pp. 39- 68.

Millar, A. & Bermudez, J., 2002. Introduction . In: J. Bermudez & A. Millar, eds. *Reasons and Nature: Essays in the theory of rationality*. New York: Oxford University Press , pp. 1-11.

Millgram, E., 2005. *Ethics Done Right: Practical Reasoning as a Foundation for Moral Theory*. Cambridge: Cambridge University Press.

Mitova, V., 2016. What Do I Care About Epistemic Norms?. In: M. Grajner & P.

Schmechtig, eds. *Epistemic Reasons, Norms and Goals*. Berlin: De Gruyter, pp. 199 - 223.

Moore, G., 1903. *Principia Ethica*. CUP 1993 ed. Cambridge: Cambridge University Press.

Nagel, T., 1985. *The View From Nowhere*. Oxford: Oxford University Press.

- Nietzsche, F., 1988. On Truth and Lies in the Nonmoral Sense. In: W. Kaufmann, ed. *The Portable Nietzsche*. New York: Penguin Books, pp. 42-48.
- Nozick, R., 1993. *The Nature of Rationality*. Princeton: Princeton University Press.
- O'Hagan, E., 2014. Shmagents, Realism and Constitutivism About Rational Norms. *Journal of Value Inquiry*, 48(1), pp. 17 -31.
- O'Neill, J., 1992. The Varieties of Intrinsic Value. *The Monist*, 75(2), pp. 119 - 137.
- Parfit, D., 1997. Reasons and Motivation. *Proceedings of the Aristotelian Society*, Volume 71, pp. 99 - 130.
- Pauer, S. H., 2018. Korsgaard's Constitutivism and the Possibility of Bad Action. *Ethical Theory and Moral Practice*, 21(1), pp. 37 - 56.
- Prinz, J., 2007. *The Emotional Construction of Morals*. Oxford: Clarendon Press.
- Putnam, H. 2003. *The Collapse of the Fact-Value Dichotomy and Other Essays*. Cambridge: Harvard University Press.
- Railton, P., 2003. On the Hypothetical and Non-hypothetical in Reasoning about Belief and Action. In: *Ethics and Practical Reason* . Oxford: Clarendon Press, pp. 53-80.
- Railton, P., 2006. Humean Theory of Practical Rationality. In: D. Copp, ed. *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press, pp. 265-281.
- Ramose, M., 2003. The Philosophy of Ubuntu and Ubuntu as Philosophy. In: P. Coetzee & A. Roux, eds. *The African Philosophy Reader*. New York: Routledge, pp. 230 - 238.



- Rawls, J., 1980. Kantian Constructivism in Moral Theory: The Dewey Lectures 1980. *Journal of Philosophy*, 77(9), pp. 303 - 358.
- Rawls, J., 1989. Themes in Kant's Moral Philosophy. In: S. Freeman, ed. *Collected Papers*. Cambridge : Harvard University Press, pp. 81 - 113.
- Rawls, J., 2000. *Lectures on the History of Moral Philosophy*. Cambridge: Harvard University Press.
- Raz, J., 1988. *The Morality of Freedom*. Oxford: Oxford University Press.
- Raz, J., 1999. *Engaging Reason*. Oxford: Oxford University Press.
- Raz, J., 2011. *From Normativity to Responsibility*. New York: Oxford University Press.
- Riccardi, M., 2018. Nietzsche's Philosophy of Mind. In: S. Lapointe, ed. *Philosophy of Mind in the Nineteenth Century*. London: Routledge, p. n.p.
- Rosati, C., 2003. Agency and the Open Question Argument. *Ethics*, 113(3), pp. 490 - 527.
- Rosati, C., 2016. Agents and "shmagents" An Essay on Agency and Normativity. In: Russ-Safer-Landau, ed. *Oxford Studies in MetaEthics 11*. Oxford: Oxford University Press, pp. 182 - 213.
- Scanlon, T., 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Scanlon, T., 2006. Reasons: A Puzzling Duality. In: R. Wallace, P. Pettit, S. Scheffler & M. Smith, eds. *Reasons and Value*. New York: Oxford University Press, pp. 231 - 246.
- Schiller, A., 2007. Psychological Nominalism and the Plausibility of Sellars. *The Southern Journal of Philosophy*, 45(3), pp. 435-454.

Schlosser, M. 2015 (revised 2019). *Agency*. Available at:

<https://plato.stanford.edu/entries/agency/> [Accessed 28 June 2017].

Schroeder, M., 2016. *Value Theory*. [Online]

Available at: <http://.plato.stanford.edu/archives/fall2016/entries/valuetheory/>

[Accessed 30 July 2018].

Schroeder, M., 2014. *Explaining the Reasons we Share*. United Kingdom: Oxford University Press.

Searle, J. 1964. How to Derive 'Ought' from "Is". *Philosophical Review*. 73 (1). Pp. 43 - 58.

Sehon, S., 2005. *Teleological Realism: Mind, Agency, and Explanation*. Cambridge: MIT Press.

Sellars, W., 1956. Empiricism and the Philosophy of Mind. *Minnesota Studies in the Philosophy of Science*, Volume 1, pp. 352-239.

Sellars, W., 1962. Philosophy and the Scientific Image of Man. In: *Science, Perception, and Reality*. London: Routledge & Kegan Paul Ltd.

Setiya, K., 2007. *Reasons without Rationalism*. Princeton: Princeton University Press.

Shafer-Landau, R., 2003. *Moral Realism: A Defense*. Oxford: Oxford University Press.

Shah, N., 2002. Clearing Space for Doxastic Voluntarism. *The Monist*, 85(3), pp. 436-445.

Shah, N., 2013. Why We Reason the Way We Do. *Philosophical Issues*, 23(1), pp. 311-325.

Shah, N. & Velleman, J., 2005. Doxastic Deliberation. *Philosophical Review*, 114(4), pp. 497-534.

Silverstein, M., 2015. The Shmagency Question. *Philosophical Studies*, 172(5), pp. 1127 - 1142.

Silverstein, M., 2016. Teleology and Normativity. In: Russ-Shafer\_Landau, ed. *Oxford Studies in Metaethics*. Oxford: Oxford University Press, pp. 214 - 240.

Simion, M., Kelp, C. & Ghijsen, H., 2016. Norms of Belief. *Philosophical Issues*, Volume 26, pp. 375 - 392.

Steglich-Petersen, A., 2006. No Norm Needed: On the Aim of Belief. *The Philosophical Quarterly*, 56(225), pp. 499-516.

Steup, M., 2017. Believing Intentionally. *Synthese*, 194(8), pp. 2673-2694.

Stocker, M., 2006. Raz on the Intelligibility of Bad Acts. In: J. Wallace, P. Pettit, S. Scheffler & M. Smith, eds. *Reason and Value*. New York: Oxford University Press, pp. 303 - 332.

Street, S., 2008. Constructivism and Reasons. *Oxford Studies in Metaethics*, Volume 3, pp. 208 - 245.

Street, S., 2010. What is Constructivism in Ethics and Metaethics. *Philosophy Compass*, 5(5), pp. 363 - 384.

Street, S., 2012. Coming to terms with contingency: humean constructivism about practical reason. In: J. Lenman & Y. Shemmer, eds. *Constructivism in practical philosophy*. Oxford : Oxford University Press, pp. 40 - 59.

Taylor, R., 1966. *Action and Prupose*. New Jersey: Prentice Hall Publishing.

Thomson, J., 1985. The Trolley Problem. *The Yale Law Journal*, 94(6), pp. 1395 - 1415.

Thomson, J., 2008. *Normativity*. Chicago: Open Court.

Tiffany, E., 2012. Why Be An Agent?. *Australasia Journal of Philosophy*, 90 (2), pp. 223 - 233.

Toppinen, T., 2015. How Norms (Might) Guide Belief. *International Journal of Philosophical Studies*, 23(3), pp. 396 - 409.

Velleman, D., 1992a. The Guise of the Good. *Nous*, 26(1), pp. 3-26.

Velleman, D., 1992b. What Happens When Someone Acts?. *Mind*, 101(403), pp. 461-481.

Velleman, D., 1996. The Possibility of Practical Reason. *Ethics*, 106(4), pp. 694-726.

Velleman, D., 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.

Velleman, D., 2004. Precis of the Possibility of Practical Reason. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 121(3), pp. 225-238.

Velleman, D., 2013. *How We Get Along*. New York: Cambridge University Press.

Wallace, R., 2001. Normativity, Commitment and Instrumental Rationality. *Philosopher's Imprint*, 1(3), p. n.p.

Wedgwood, R., 2002. The Aim of Belief. *Philosophical Perspectives*, 36(16), pp. 267 - 297.

Wedgwood, R., 2007. *The Nature of Normativity*. Oxford: Clarendon Press.

Wedgwood, R. Must Rational Intentions Maximise Utility? *Philosophical Explorations*, 20:sup2, pp. 73- 92

Wegner, D.M., 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Westermarck, E., 1960. *Ethical Relativity*. Paterson, NJ: Littlefield, Adams.

- Wiggins, D., 1987. A Sensible Subjectivism. In: *Needs, Values, Truth: Essays in the Philosophy of Value*. Oxford: Blackwell.
- Williams, B., 1973. *Deciding to Believe. Problems of the Self*. Cambridge: Cambridge University Press.
- Williams, B., 1979. Internal and External Reasons. In: R. Harrison, ed. *Rational Action*. Cambridge: Cambridge University Press, pp. 101 - 113.
- Williams, B., 1981. Persons, Character, and Morality. In: *Moral Luck*. New York: Cambridge University Press.
- Williams, B., 1985. *Ethics and the Limits of Philosophy*. Cambridge: Harvard University Press.
- Williamson, T., 2000. *Knowledge and its limits*. Oxford: Oxford University Press.
- Wimpenny, J. et al., 2009. Cognitive Processes Associated with Sequential Tool Use in New Caledonian Crows. *PlosOne*, 4(8), pp. 1-16.
- Wong, D., 1984. *Moral Relativity*. Berkeley CA: University of California Press.
- Wong, D., 2006. *Natural Moralities: A Defense of Pluralistic Relativism*. New York: Oxford University Press.
- Wright, C., 2002. On Basic Logical Knowledge: Reflections on Paul Boghossian's 'How are Objective Epistemic Reasons Possible?'. In: J. Bermudez & A. Millar, eds. *Reasons and Nature: Essays in the Theory of Rationality*. New York: Oxford University Press, pp. 49 - 84.
- Wright, C., 2012. Comment on Paul Boghossian, "The Nature of Inference". *Philosophical Studies*, Issue doi: 10.1007/s11098-012-9892-9.

Ylikoski, P.K., 2013. Causal and Constitutive Explanation Compared. *Erkenntnis*, 78 (2). pp. 277-297

