What We Informationally Owe Each Other

Chapter 4 in *Algorithms & Autonomy: The Ethics of Automated Decision Systems* (Cambridge University Press, forthcoming)

Alan Rubel
Information School and Center for Law, Society & Justice
University of Wisconsin, Madison

Clinton Castro
Department of Philosophy
Florida International University

Adam Pham
Division of Humanities and Social Science
California Institute of Technology

4. What we informationally owe each other

Abstract: One important criticism of algorithmic systems is that they lack transparency. Such systems can be opaque because they are complex, protected by patent or trade secret, or deliberately obscure. In the EU, there is a debate about whether the General Data Protection Regulation (GDPR) contains a "right to explanation," and if so what such a right entails. Our task in this chapter is to address this informational component of algorithmic systems. We argue that information access is integral for respecting autonomy, and transparency policies should be tailored to advance autonomy.

To make this argument we distinguish two facets of agency (i.e., capacity to act). The first is *practical* agency, or the ability to act effectively according to one's values. The second is what we call *cognitive* agency, which is the ability to exercise what Pamela Hieronymi calls "evaluative control" (i.e., the ability to control our affective states, such as beliefs, desires, and attitudes). We argue that respecting autonomy requires providing persons sufficient information to exercise evaluative control and properly interpret the world and one's place in it. We draw this distinction out by considering algorithmic systems used in background checks, and we apply the view to key cases involving risk assessment in criminal justice decisions and K-12 teacher evaluation.

In chapter 2, we articulated our conception of autonomy. We argued for a lightweight, ecumenical approach that encompasses both psychological and personal autonomy. In chapter 3, we drew on this account to set out conditions that are crucial in determining whether algorithmic decision systems respect persons' autonomy. Specifically, we argued that algorithmic decision systems are justifiable to the extent that people subject to them can reasonably endorse them. Whether people can reasonably endorse those systems turns on conditions of reliability, responsibility, stakes, and relative burden.

Notice, though, that the conditions set out in chapter 3 are primarily about how those systems threaten persons' material conditions, such as whether teachers are fired based on evaluation systems and whether defendants are subject to more stringent conditions based on risk assessment systems. But people are not just passive subjects of algorithmic systems—or at least they ought not to be—and whether use of a system is justifiable overall turns on more than the material consequences of its use.

In this chapter we will argue that there is a distinct *informational* component to respecting autonomy. Specifically, we owe people certain kinds of information and informational control. To get a basic sense of why, consider our understanding of autonomy from chapter two, which has two broad facets. Psychological autonomy includes conditions of competence (including epistemic competence) and authenticity. Personal autonomy includes procedural and substantive independence, which at root demand space and support for a person to think, plan, and operate. Further, as we explain in chapter 2, whether agents are personally autonomous turns on the extent to which they are capable of incorporating their values into important facets of their lives. Respecting an agent's autonomy requires that one not deny her what she needs to incorporate her values into important facets of her life. It is a failure of respect to prevent agents from exercising their autonomy, and it is wrongful to do so without sufficiently good reason. Incorporating one's values into important facets of one's life requires that one have access to relevant information. That is, autonomy requires having information important to one's life, and respecting autonomy requires not denying agents that information (and at times making it available). Algorithmic decision systems are often built in a way that prevents people from understanding their operations.[1] This may, at least under certain circumstances, preclude persons' access to information to which they have a right.

[1] Frank Pasquale (2016) argues that lack of transparency is one of the defining features and key concerns of technological "black boxes" that exert control over large swaths of contemporary life. Such obscurity can derive from many sources, including technological complexity, legal protections via intellectual property, and deliberate

That is the broad contour of our argument. Our task in the rest of the chapter is to fill that argument in. We'll begin by describing two new cases, each involving background checks, and analyzing those cases using the Reasonable Endorsement Test we develop in chapter 3. We then explain important facets of autonomy that are missing from the analysis. To address that gap, we distinguish several different modes of agency, including *practical* and *cognitive* agency. We argue that individuals have rights to information about algorithmic systems in virtue of their practical and cognitive agency. Next, we draw on some scholarship surrounding a so-called "right to explanation" in the European Union's General Data Protection Regulation and how those relate to our understanding of cognitive and practical agency. Finally, we then apply our criteria to our polestar cases.

To be clear, we are not arguing that individuals have a right to all information that is important in understanding their lives, incorporating their values into important decisions, and exercising agency. Rather, we argue that they have some kind of defeasible claim to such information. Our task here is to explain the basis for that claim, the conditions under which it creates obligations on others to respect, and the types of information the moral claims underwrite. A recent report on ethics in AI systems states, "Emphasis on algorithmic transparency assumes that some kind of 'explainability' is important to all kinds of people, but there has been very little attempt to build up evidence on which kinds of explanations are

obfuscation. For our purposes the source of obscurity is initially less important than what autonomy demands. The source will become important when evaluating what duties people have to provide information as a matter of respecting others' autonomy.

desirable to which people in which contexts."[2] We hope to contribute to this issue with an argument about what information is warranted.

4.1. The misfortunes of Catherine Taylor and Carmen Arroyo

Let's begin by considering two new cases.

Arkansas resident Catherine Taylor was denied a job at the Red Cross. Her rejection letter came with a nasty surprise. Her criminal background report included a criminal charge for intent to manufacture and sell methamphetamines.[3] But Taylor had no criminal history. The system had confused her with *Illinois* resident Catherine Taylor, who had been charged with intent to manufacture and sell methamphetamines.[4]

Arkansas Catherine Taylor wound up with a false criminal charge on her report because ChoicePoint (now a part of LexisNexis), the company providing the report, relied on bulk data to produce an "instant" result when checking her background.[5] This is a common practice. Background screening companies such as ChoicePoint generate reports through automated processes that run searches through large databases of aggregated data, with minimal (if any) manual overview or quality control. ChoicePoint actually had enough accurate information—such as Taylor's address, Social Security number, and credit report—to avoid tarnishing her

[2] Jess Whittlestone et al., "Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research" (London: The Nuffield Foundation, 2019), 12.

[3] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, 1st ed. (New York: Crown, 2016).

[4] Persis S. Yu and Sharon M. Dietrich, "Broken Records: How Errors by Criminal Background Checking Companies Harm Workers and Businesses" (National Consumer Law Center, April 11, 2012).

[5] Yu and Dietrich.

reputation with mistakes.[6] Unfortunately for Taylor, the product ChoicePoint used in her case simply wasn't designed to access that information.[7]

ChoicePoint compounded the failure by refusing to rectify its mistake. The company said it could not alter the sources from which it draws data. So, if another business requested an "instant" report on Arkansas Catherine Taylor, the report would include information on Illinois Catherine Taylor.[8]

This is not the only occasion on which Arkansas Taylor (of Arkansas) would suffer this kind of error. Soon after learning about the ChoicePoint mix-up, she found at least ten other companies who were providing inaccurate reports about her. One of those companies, Tenant Tracker, conducted a criminal background check for Taylor's application for federal housing assistance that was even worse than ChoicePoint's check. Tenant Tracker included the charges against Illinois Catherine Taylor and *also* included a separate set of charges against a person with a different name, Chantel Taylor (of Florida).[9]

Taylor's case is not special. Another background screening case involving a slightly different technology shows similar problems. It is common for background screeners to offer products that go beyond providing raw information on a subject and produce an algorithmically generated judgement in the form of a score or some other kind of recommendation. "CrimSAFE," which was developed by CoreLogic Rental Property Solutions, LLC (CoreLogic),

[6] Yu and Dietrich.

[7] Yu and Dietrich citing Deposition of Teresa Preg at 63-64.

[8] Yu and Dietrich.

[9] O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*.

is one such product.[10] CrimSAFE is used to screen tenants. CoreLogic markets it as an "automated tool" that "processes and interprets criminal records and notifies leasing staff when criminal records are found that do not meet the criteria you establish for your community."[11]

When a landlord or property manager uses CrimSAFE to screen a tenant, CoreLogic delivers a report that indicates whether CrimSAFE has turned up any disqualifying records.[12] But the report does not indicate what those allegedly disqualifying records are or any information about them (such as their dates, natures, or outcomes). To reiterate, the report only states *whether* disqualifying records have been found, not *what* they are. CoreLogic provides neither the purchaser nor the subject of the report any of the underlying details.[13]

Let's now look at a particular case involving CrimSAFE. In July of 2015, Carmen Arroyo's son Mikhail suffered an accident that left him unable to speak, walk, or care for himself.[14] Carmen was Mikhail's primary caregiver, and she wanted to have Mikhail move in with her when he was discharged from treatment. For Mikhail to move into his mother's apartment, he had to be screened by her complex, and so the complex manager had CoreLogic screen Mikhail using CrimSAFE.[15]

[10] Ariel Nelson, "Broken Records Redux: How Errors by Criminal Background Check Companies Continue to Harm Consumers Seeking Jobs and Housing" (Boston, MA: National Consumer Law Center, December 6, 2019).

[11] Nelson. See also Connecticut Fair Hous. Ctr. v. Corelogic Rental Prop. Sols., LLC, 369 F. Supp. 3d 362 (D. Conn. 2020).

[12] Nelson, "Broken Records Redux: How Errors by Criminal Background Check Companies Continue to Harm Consumers Seeking Jobs and Housing."

[13] Nelson.

[14] Nelson.

[15] Nelson.

CoreLogic returned a report to the apartment complex manager indicating that Mikhail was not fit for tenancy, based on his criminal record.[16] The report did not specify the date, nature, or outcome of any criminal charges on Mikhail's record. Further, Mikhail had never been convicted of a crime. Despite being unaware of the date, nature, or outcome of the alleged criminal conduct—and without taking into consideration the question of whether Mikhail was at that point even capable of committing the crimes he had been accused of—the manager adopted CoreLogic's conclusion and denied Mikhail tenancy.[17] Hence, Carmen Arroyo was unable to move her severely injured son into her apartment where she could provide the care he needed.

Taylor and the Arroyos have suffered serious harms. And knowing the causes of their misfortunes is of little help in reversing those misfortunes. Decisions based on faulty criminal background reports are rarely overturned after those reports are identified as faulty.[18] As the National Consumer Law Center puts it, "you can't unring the bell."[19]

Taylor learned of the problems with her background as her tribulations unfolded. Arroyo learned of the problem only after being denied the key thing she needed to support her son, though she did eventually learn the reasons for Mikhail being denied tenancy. Many who are denied housing or employment through automated screening do not ever learn why.[20]

[16] Nelson.

[17] Nelson.

[18] Yu and Dietrich, "Broken Records: How Errors by Criminal Background Checking Companies Harm Workers and Businesses."

[19] Yu and Dietrich.

[20] For further discussion of background check algorithms and lack of regulation and oversight, see Lauren Kirchner and Matthew Goldstein, "Access Denied: Faulty Automated Background Checks Freeze Out Renters," May 28, 2020, https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters.

One reason people do not find out is that under U.S. law, consumer reporting agencies (companies that provide reports on consumers, such as background checks) do not have to tell the subjects of background checks that they are being screened. The relevant statute in this context is the Fair Credit Reporting Act (FCRA), which requires either notification *or* the maintenance of strict procedures to ensure that the information is complete and up to date.[21] This leaves reporting agencies the legal option of leaving the subjects of background searches out of the loop.

Further, many companies that provide background checks maintain that they are not consumer reporting agencies at all. So, they maintain that the FCRA does not apply to them. As a result, they neither notify subjects of background checks nor maintain the strict procedures necessary to ensure the information in their systems is complete and up to date. One of the companies responsible for disseminating false information about Catherine Taylor, PublicData.com, simply denies that it is a consumer reporting agency.[22] When Taylor notified PublicData.com of the errors it had made about her, they were unwilling to do anything to correct the errors.[23] This was a matter of company policy, which is explicit that it “will NOT modify records in any database upon notification of inaccuracies.”[24]

[21] 91st United States Congress, “An Act to Amend the Federal Deposit Insurance Act to Require Insured Banks to Maintain Certain Records, to Require That Certain Transactions in U.S. Currency Be Reported to the Department of the Treasury, and for Other Purposes.,” 15 U.S.C. § 1681 § (1970); Yu and Dietrich, “Broken Records: How Errors by Criminal Background Checking Companies Harm Workers and Businesses.”

[22] Yu and Dietrich, “Broken Records: How Errors by Criminal Background Checking Companies Harm Workers and Businesses.”

[23] Yu and Dietrich.

[24] Yu and Dietrich.

FCRA also requires employers using background checks to disclose that they will be doing background checks and to notify a candidate if adverse action may be taken in response to a background check.[25] However, employers often do not comply with notice requirements.[26]

4.1.1. Taylor, Arroyo, and the Reasonable Endorsement Test

One way to understand Taylor's and the Arroyos' situations is in the terms we spell out in chapter 3, namely whether the background reporting systems are ones that people subject to them can reasonably endorse. Both Taylor and Arroyo have experienced considerable material burdens based on algorithmically aided decision systems. Both were held to account by systems that are based on factors for which Taylor and Arroyo are not responsible, and the stakes in each case are high. Hence, one could make the case that the reporting systems are ones that individuals subject to them cannot reasonably endorse as comporting with their material interests. Such an analysis, while compelling, would not be complete.

Something has gone wrong in the Taylor and Arroyo cases beyond the fact that they were materially harmed. This separate consideration is an informational wrong. Taylor and Arroyo did not know (at least initially) what information in their files led to their background check results. Arroyo did not discover the basis for Mikhail's check until it was too late to do anything meaningful about it. Taylor lost opportunities before she discovered the reason. Further, in Taylor's case, several companies providing the misinformation would not fix their files upon

[25] 91st United States Congress, An Act to amend the Federal Deposit Insurance Act to require insured banks to maintain certain records, to require that certain transactions in U.S. currency be reported to the Department of the Treasury, and for other purposes.

[26] Yu and Dietrich, "Broken Records: How Errors by Criminal Background Checking Companies Harm Workers and Businesses."

learning that they had made a mistake. Finally, both Taylor and Arroyo were in left in the dark as to how, exactly, the results came out the way they did; they were not afforded an understanding of the systems that cost them the opportunities they had sought.

Arroyo has an additional, distinctive complaint. When her son's application was rejected, the apartment complex did not know the details of the disqualifying conduct because CoreLogic did not supply them. This means that Arroyo was not given enough information about Mikhail's rejection to even contest the claim. Compare Arroyo's case with Taylor's. Taylor at least knew that her file had contained a false drug charge. Knowing what she had been accused of informed her that she had to prove what she hadn't done. Arroyo lacked even that.

We have mentioned that there is at least some regulation that attempts to address these sorts of issues and that there is plausibly a question as to whether CoreLogic complies with its legal obligations under FCRA (as stated above, companies do not always follow the notification requirement). Could full compliance with FCRA bring about practices that Taylor and Arroyo could reasonably endorse? Again, we think not. For one, FCRA does not specify when subjects are owed notification.[27] So, the notification requirement can be met without actually affording data subjects the underlying thing that really matters: time to effectively respond to any false or misleading information in their files and an understanding of where they stand with respect to decisions made about them. These are the claims we address in the following section.

[27] Yu and Dietrich.

4.2. Two arguments for informational rights

Surely the Taylor and Arroyo cases grate on our intuitions, both because of the harms resulting from their background checks and the fact that each was in the dark about those checks. Such intuitions, however, can only take us so far. We need an argument to explain the wrongs adequately. Our argument is that persons' autonomy interests have a substantial informational component that is distinct from the material components we argue for in chapter 3. Specifically, respecting the autonomy of persons subject to algorithmic decision systems requires ensuring that they have a degree of cognitive access to information about those systems.

Agency refers to action and the relationship between a person (or other entity) and actions that are in some sense attributable to that person. That relationship may be merely causal (as when a person hands over their wallet at gunpoint), it may be freely willed, it may be deliberately planned, or it may be something else. Hence, agency is broader than autonomy, for a person may be an agent but neither psychologically nor personally autonomous. However, agency is morally important in that persons have claims to exercise agency (and to have room to exercise agency) in light of their (capacity) autonomy. On the relationship between autonomy and agency, Oshana writes: "An autonomous person is an agent—one who directs or determines the course of her own life and who is positioned to assume the costs and the benefits of her choices."[28] We return to the relationship between agency and autonomy, and the relation of both to conceptions of freedom, in chapters 5 and 6.

[28] Marina Oshana, *Personal Autonomy in Society*, 1st ed. (Aldershot, Hants, England ; Burlington, VT: Routledge, 2006), vii.

To make our case, we first need to distinguish two aspects of agency. At base, agency is the capacity (or effective exercise of the capacity) to act. And agents are beings with such capacity.[29] There is substantial philosophical controversy surrounding conceptions and metaphysics of agency (e.g., whether it is simply a causal relation between an actor and event, whether agency requires intentionality, the degree to which non-humans may be agents). We can leave many of those to the side so that we can focus on agency with respect to action and mental states.

The most familiar facet of agency is the ability to act physically in a relatively straightforward way, for example taking a walk, preparing a meal, or writing an email. A more complex exercise of agency involves taking actions that institute a plan or that realize one's values (which is to say, exercise agency in such a way that doing so successfully instantiates one's psychological autonomy). Call this "practical agency." Exercising practical agency so that it is consistent with one's preferences and values requires a great deal of information and understanding. So, for example, if it's important to a person to build a successful career, then it is important for her to understand how her profession and her organization function, how to get to work, how to actually perform tasks assigned, and so forth. And if that person's supervisor fails to make available information that is relevant to her job performance, the supervisor fails to respect the person's practical agency because doing so creates a barrier to the employee incorporating her values into an important facet of her life. Notice that this understanding of practical agency shares similar foundations to the substantive independence requirement of

[29] Sven Nyholm puts it like this: "agency is a multidimensional concept that refers to the capacities and activities most centrally related to performing actions, making decisions, and taking responsibility for what we do." Sven Nyholm, *Humans and Robots: Ethics, Agency, and Anthropomorphism* (London; New York: Rowman & Littlefield Publishers, 2020), 31.

personal autonomy outlined in chapter 2. Being denied important information about the practicalities of planning and living one's life undermines the degree to which one has substantive independence from others.

The importance of information to exercising agency does not solely depend on agents' abilities to use information to guide actions. A second aspect of agency is the ability to *understand* important facets of one's life. Call this "cognitive agency." The distinction between practical agency and cognitive agency tracks Pamela Hieronymi's view that ordinary intentional agency, in which we exercise control over actions—deciding to take a walk, deciding to prepare a meal—is distinct from "mental agency" (although we use 'cognitive agency,' the notion is the same). Mental agency, Hieronymi explains, is the capacity to exercise *evaluative* control over our mental states (e.g., our attitudes, beliefs, desires, and reactive responses). The difference between ordinary intentional agency and mental agency is the difference between an actor deciding "whether to do" (i.e., whether to take some action in the world beyond oneself) and the actor deciding "whether to believe." Hieronymi's view is that agents indeed exercise control—to some degree, and within important limits—over how they respond mentally to their circumstances. The scope of one's evaluative control over one's mental states and the extent to which one can exercise it effectively are less important to our project than recognizing the domain of cognitive agency.[30]

Cognitive agency grounds moral claims in much the same way as practical agency. Respecting persons as autonomous requires that they be able to incorporate their sense of value

[30] For a similar division of aspects of our agency and discussion, see Michael Smith, "A Constitutivist Theory of Reasons: Its Promise and Parts," *Law, Ethics and Philosophy* 1 (December 1, 2013): 9–30.

into decisions about conducting their lives as a matter of practical agency. Similarly, respecting persons as autonomous requires that they be able to incorporate their sense of value into how they understand the world and their place in it. As Thomas Hill, Jr. has argued, deception is an affront to autonomy regardless of whether that deception changes how one acts because it prevents persons from properly interpreting the world; even a benevolent lie that spares another's feelings can be an affront because it thwarts that person's ability to understand her situation.[31] We can extend Hill's argument beyond active deception. Denying agents information relevant to important facets of their lives can circumvent their ability to understand their situation just as much as deceit.[32] In other words, deceit circumvents persons' epistemic competence and may render their desires and beliefs inauthentic.

One might question here whether practical and cognitive agency are distinctive issues for algorithmic systems. Strictly speaking, the answer is no, because—as we explain in chapter 1—many of the arguments we advance in this book are applicable to a wide range of social and technical systems. However, there are several reasons to think that practical and cognitive agency raise issues worth analyzing in the context of algorithmic systems. For one, humans are well-adapted to understanding, regulating, and interacting with other humans and human systems, but the same is not true of artificial systems. Sven Nyholm has recently argued that there are a number of important moral issues that arise in the context human-robot interactions precisely because humans tend to attribute human-like features to robots, when in fact humans

[31] Thomas Hill, Jr., "Autonomy and Benevolent Lies," *The Journal of Value Inquiry* 18, no. 4 (December 1, 1984): 251–67.

[32] Alan Rubel, "Privacy and the USA Patriot Act: Rights, the Value of Rights, and Autonomy," *Law and Philosophy* 26, no. 2 (2007): 119–159.

have a poor grasp of what robots are like.[33] The same can be said for algorithmic systems. Related is that the informational component of algorithmic systems may be more pronounced than it is for bureaucratic or other primarily human decisions. We may understand the limited, often arbitrary nature of human decisions. But infirmities of algorithmic systems may be harder for us to reckon, and we may lack the kinds of heuristics we can employ to understand human decision-making.

The view so far is that information is important for practical agency and cognitive agency, and that claims to such information are grounded in autonomy. Surely, however, it isn't the case that respecting autonomy requires providing *any* sort of information that happens to advance practical and cognitive agency. After all, some information may be difficult to provide, may be only modestly useful in fostering agency, or may undermine other kinds of interests. Moreover, some information may be important for exercising practical and cognitive agency, but no one has an obligation to provide it. If one wants to feel better by cooking healthier meals, information about ingredients, recipes, and techniques is important in exercising practical agency over one's eating habits. However, it is not clear that anyone thwarts another person's agency by failing to provide that information. What we need, then, is a set of criteria for determining if and when informational interests are substantial enough that persons have claims to that information on the grounds of practical agency or cognitive agency.

[33] Nyholm, *Humans and Robots*, 15–18.

4.2.1. Argument 1: practical agency

The first set of criteria for determining whether persons have claims to information about automated decision systems echoes the criteria we advance in chapter 3. Specifically, whether an individual has a claim to information about some algorithmic decision system that affects their life will be function of that system's reliability, the degree to which it tracks actions for which they are responsible, and the stakes of the decision.

Assume for a moment that Taylor's problems happen in the context of a reporting system that people can't reasonably reject on grounds of reliability, responsibility, and stakes. Taylor nonetheless has a claim based on *practical agency*. To effectively cope with the loss of her opportunities for employment and credit, she needs to understand the source of her negative reports. To that extent, Taylor's claims to information based on practical agency resemble those of anyone who is subject to credit reports and background checks. And, of course, Taylor did indeed have access to very general information about the nature of background checks and credit reporting. That might have been sufficient to understand that her background check was a factor in her lost opportunity.

We can capture this sense of Taylor's claims with what we'll call the

> **Principle of Informed Practical Agency (PIPA):** One has a defeasible claim to information about decision systems affecting one's life where (a) that information advances practical agency, (b) it advances practical agency because one's practical agency has been restricted by the operations of that system, (c) the effects of the decision system bear heavily on significant facets of one's life,

> and (d) information about the decision system allows one to correct *or* mitigate its effects.

Surely this principle holds, but it cannot capture the degree to which Taylor’s practical agency was thwarted by ChoicePoint and other reporting agencies. Rather, a key limitation on Taylor’s practical agency is the fact that the reporting agencies systemically included misinformation in her reports. In other words, Taylor’s claims to information are particularly weighty because the background checks at once purport to be grounded in information for which she is responsible (including criminal conduct) *and* the reports were systemically wrong. Hence, to capture the strength of Taylor’s claims, we can add the following:

> **Strong Principle of Informed Practical Agency:** a person’s claim to information based on the PIPA is stronger in case (e) the system purports to be based on factors for which a person is responsible, and (f) the system has errors (even if not so frequent that they, on their own, make it unendorseable).

Knowing that the background checking system conflates the identities of people with similar names, knowing that her own record includes information pertaining to other people with criminal records, and knowing that the system relies on other background checking companies’ databases and thus re-populates her profile with mistaken information can provide Taylor with tools to address those mistakes. That is, she can better address the wrongs that have been visited upon her by having information about the system that makes those wrongs possible. To be clear, a greater flow of information to Taylor does not make the mistakes and harms to her any less wrongful. Even if it is true that a system is otherwise justifiable, respecting autonomy demands support for practical agency so that people may address the infirmities of that system.

What is key for understanding claims based on practical agency is the distinction we make in chapter 2 between local autonomy (the ability to make decisions about relatively narrow facets of one's life according to one's values and preferences) and global autonomy (the ability to structure larger facets of one's life according to one's values and preferences). In many contexts, respect for autonomy is local. Informed consent for undergoing a medical procedure, participating as a subject in research, agreeing to licensing agreements, and the like have to do with whether a person can act in a narrow set of circumstances. Our principles of practical agency, in contrast, concern aspects of autonomy that are comparatively global. One rarely (if ever) provides meaningful consent to having one's data collected, shared, and analyzed for the purposes of background checks, and hence enjoys only a little local autonomy over that process.[34]

Individuals have little (if any) power to avoid credit and background checks, and hence do not have global autonomy with respect to how they are treated. However, understanding how their information is used, whether there is incorrect information incorporated into background checks, and how that incorrect information precludes them from opportunities may be important (as in Taylor's case) in order to prevent lack of local autonomy from becoming relatively more global. That is, mitigating the effects of algorithmic systems may allow one to claw back a

[34] Daniel J. Solove, "Privacy Self-Management and the Consent Dilemma," *Harvard Law Review* 126 (2013): 1880–1903.

degree of global autonomy. And that ability to potentially exercise more global autonomy underwrites a moral claim to information.

The two principles of informed practical agency only tell us so much. They cannot, for example, tell us precisely what information one needs. In Taylor's case, practical agency requires understanding something about how the algorithmic systems deployed by ChoicePoint actually function, who uses them for what purposes, and how they absorb information (including false information) from a range of sources over which they exercise no control and minimal (if any) oversight. But other decision systems and other circumstances might require different kinds of information. The principles also cannot tell us exactly *who* needs to be afforded information. While the claim to information in this case is Taylor's, it may be that her advocate, representative, fiduciary, or someone else should be the one who actually receives or accesses the relevant information. Taylor, for instance, might have a claim that her employer learn about the infirmities in ChoicePoint and Tenant Tracker's algorithmic systems. The principles cannot tell us the conditions under which persons' claims may be overridden.

The principles discussed so far only address the epistemic side of practical agency. But Taylor is owed more than just information. We can see this by considering one of the most deeply troubling facets of her case: the reluctance that the data controllers who are involved have toward fixing her mistaken data. One effect of their reluctance is that it undercuts her ability to realize her values, something to which she has a legitimate claim. To capture this, we need—in addition to the principles of informed agency—a principle that lays bare agents' claim to control:

> **Principle of Informational Control:** One has a defeasible claim to make corrections to false information fed into decision systems

> affecting one's life where (a) one's practical agency has been restricted by the operations of that system, (b) the effects of the decision system bear heavily on significant facets of a person's life, and (c) correcting information about the decision system allows one to correct *or* mitigate its effects.

As before, we need a second principle specifying certain cases where this claim is stronger:

> **Strong Principle of Informational Control:** a person's claim to correct information based on the PIC is stronger in case the system purports to be based on factors for which a person is responsible.

These principles demand of the systems used in the Taylor's case that she not only be able to learn what information a system is based on, but that she be able to contest that information when it is inaccurate. The claim she has in this case is (just like the principles of informed practical agency) grounded in her agency, i.e., her claim to decide what is valuable for herself and pursue those values so long as they are compatible with respect for the agency and autonomy of others.

Now, the principles of informed practical agency and informational control cannot tell us what a person's informational claims are in cases where they are unable to exercise practical agency. We consider that next.

4.2.2. Argument 2: cognitive agency

Cognitive agency can also ground a claim to information. Consider a difference between the Taylor and Arroyo cases. Or, more specifically, a difference between Taylor's case once she had experienced several iterations of problems with her background checks and Arroyo's case after she'd been denied housing with her son. Taylor at some point became aware of a system that treats her poorly and for which she bears no responsibility. Arroyo, in contrast, was precluded from moving her son into her apartment for reasons she was unable to ascertain, the basis for the decision was an error, and the result was odious. Denying tenancy to Arroyo's son Mikhail is surely an injustice. But that wrong is compounded by its obscurity, which precluded Arroyo from interpreting it properly. That obscurity violates what we'll call the:

> **Principle of Informed Cognitive Agency:** One has a defeasible claim to information about decision systems affecting significant facets of a person's life (i.e., where the stakes are high).

As before—and for familiar reasons—we will add a second, stronger principle:

> **Strong Principle of Informed Cognitive Agency:** a person's claim to information based on the PICA is stronger in case the system purports to be based on factors for which a person is responsible.

Arroyo is an agent capable of deciding for herself how to interpret the decision, and she deserves the opportunity to do so. Her ability to understand her situation is integral in her exercising cognitive agency, but the facts that are crucial for her understanding are that her ability to care for her son is a function of the vagaries of a background check system.

Cognitive agency is implicated in Arroyo's case in part because her predicament is based on a system that bears on an important facet of her life (being able to secure a place to live and care for one's child) and purports to be based on actions for which she is responsible (criminal conduct). The system, meanwhile, is such that it treats old charges as dispositive even though they were withdrawn, and as remaining dispositive regardless of whether the person is at present in any position to commit such a crime at all. The reason such facts about the background check system are important is not because they will allow Arroyo to act more effectively to mitigate its effects. She was unable to act effectively when she was precluded from moving her son into her apartment. Rather, those facts are important for Arroyo to be able to act as a cognitive agent by exercising evaluative control over what to believe and how to interpret the incident.

Notice that the criteria for a claim to information based on cognitive agency appears less stringent than for practical agency. However, it does not follow that cognitive agency demands more information. Rather, cognitive agency demands *different kinds* of information. Because practical agency requires information sufficient to effectively act, it may require technical or operational information. Cognitive agency, in contrast, requires only enough information to exercise evaluative control. In the context of background checks, this might require only that one be able to learn that there is an algorithmic system underlying one's score, that the system has important limitations, that it is relatively unregulated (as, say, compared to FICO credit score reporting), and the factors that are salient in determining outcomes.[35]

Of course, that leaves us with the question of what information is necessary to exercise evaluative control. Our answer is whatever information is most morally salient, and the claim to information increases as the moral salience of information increases. So, in the case of Arroyo's

[35] See also the discussion of counterfactual explanations in section 4.4.2.

background check, morally salient information includes the fact of an automated system conducting the check and the fact that her son's current condition did not enter the assessment. It is true that there might be other morally salient information. For example, we can imagine a case where the future business plans of CoreLogic is peripherally morally salient to a case; however, a claim to that information is comparatively weaker, and hence more easily counterbalanced by claims CoreLogic has to privacy in its plans.

4.2.3. Objections and democratic agency

There are a several objections to the view we have set out so far that are important to address here. The first is that it proves too much. There are myriad and expanding ways that algorithmic systems affect our lives, and information about those systems bears upon our practical and cognitive agency in innumerable ways. Hence, the potential scope for individuals' claims to information is vast.

It is certainly true that the principles of informed practical agency and of informed cognitive agency are expansive. However, the principles have limitations that prevent them from justifying just any old claim to information. To begin, the principles of practical agency require that an algorithmic system restrict an individual's practical agency. How to determine what counts as a restriction, of course, is an interpretative difficulty. For example, does an algorithmic system that calculates one's insurance premiums restrict one's practical agency? What about a system that sets the prices one is quoted for airline tickets? Nonetheless, even on a capacious interpretation, it won't be just *any* algorithmic system that affects one's practical agency. Another significant hurdle is that the algorithmic system must affect significant facets of a person's life. Perhaps insurance rates and airline prices clear that hurdle, but it is close. Other systems, such as what political ads one is served in election season, what music is recommended on Spotify, or which route Google maps suggests to your destination do not impose restrictions on one's practical agency.[36] The requirement that information allow a person to correct or mitigate the effects of an algorithmic system, therefore, is a substantial hurdle for the information to clear. Claims to information that have no such effect would fall under cognitive agency (and as we explain below, information that respects cognitive agency is less onerous to provide).

A second, related, objection is that many people—probably most people—will not wish to use information to exercise practical or cognitive agency. It is cheap, so to speak, to posit a claim to information, but it is pricey for those who deploy algorithmic systems, and the actual payoff is limited. This criticism is true so far as it goes, but it is compatible with the principles we've offered. For one, the fact (if it is) that many people will not exercise practical agency does not say much in itself about the autonomy interests one might have in a piece of information. This is much the same as in the case of medical procedures: few people opt out of care, but information about care remains necessary to respecting their autonomy interests. Moreover, the objection speaks mostly to the *strength* of individuals' claims. All else equal, the higher the stakes involved, and the more information can advance practical agency, the stronger the claims. And the more unwieldy it is for entities using algorithmic systems to provide information, the greater are countervailing considerations.

[36] There is a related question about the baseline against which some action counts as a restriction. A direction-suggesting algorithm (e.g., Google Maps) in most cases increases one's practical agency by allowing one to find one's way quickly and easily. In the rare case that such a system sends one on a sub-optimal route, we could interpret that as a restriction of practical agency against a baseline of an overall expansion of practical agency. The best understanding of the principles of practical agency, though, is against a baseline of no algorithmic system.

A third objection is that the arguments prove too *little*. There is presumably a lot of information to which people have some sort of claim, but which does not advance individuals' practical or cognitive agency. To introduce this objection, let's start with a claim to information based on cognitive agency. Imagine a person (call him DJ) born into enormous advantage: wealth, social status, educational opportunities, political influence, and so forth. Suppose, however, that these advantages derive almost entirely from a range of execrable practices by DJ's family and associates: child labor, knowingly inducing addiction to substances that harm individuals and hollow out communities, environmental degradation, and so forth. DJ's parents, we might imagine, shield him from the sources of his advantage as he grows up, and when he reaches adulthood, he does not inherit any wealth (though of course he retains all the social, educational, and political benefits of his privileged upbringing). The degree to which his ignorance limits his practical agency is not clear, given his advantages.[37] However, on the view we outline in the previous section, DJ's parents certainly limit his cognitive agency by continuing to shield him from the sources of his advantage; he is precluded from understanding important facts about his life, as well as the chance to interpret his circumstances in light of those facts.

DJ is not the only person whose cognitive agency is a function of understanding the source of enormous wealth and advantage. Anyone who has an interest in their society's social, political, financial, and educational circumstances has some claim to understand how DJ's family's and associates' actions bear upon those circumstances. And that is true regardless of

[37] To the extent that DJ wishes to steer his course on the basis his family and social background and reconcile that with his values and beliefs, shielding him may indeed limit his practical agency.

whether they are in any position to change things. In other words, it is the fact that DJ's family's actions have an important effect on the world that grounds others' claims to information, not strictly how those actions affect each individual.[38] But it is difficult to see how the importance of that information is a function of either practical or cognitive agency.

With that in mind, let's return to algorithmic systems. In path-breaking work, Latanya Sweeney examined Google's AdSense algorithm, which served different advertisements, and different *types* of advertisements, based on names of search subjects.[39] Sweeney's project began with the observation that some advertisements appearing on pages of Google search results for individuals' names suggested that the individuals had arrest records. The project revealed that the ads suggesting arrest records were more or less likely to appear based on whether a name used in the search was associated with a racial group. That is, advertisements suggesting arrest records appeared to show up more often in Google ads served for searches that included names associated with Black people than in ads served for searches that included names associated with White people. This result was independent of whether the searched names actually had arrest records.[40] While Sweeney did not have access to the precise mechanism by which the AdSense algorithm learned to serve on the basis of race, as she explains, a machine learning system could achieve this result over time simply by some number of people clicking on ads suggesting arrest records that show up when they use Google to search for Black identifying names.[41]

[38] There might be plausible rationales for continued secrecy, e.g., privacy rights. But those are countervailing considerations to individuals' autonomy interests—in this case grounded in cognitive agency.

[39] Latanya Sweeney, "Discrimination in Online Ad Delivery," *Communications of the ACM* 56, no. 5 (January 28, 2013): 44–54.

[40] Sweeney.

[41] Results from algorithmic systems that differ on the basis of race and ethnicity are rampant. Examples include predominantly sexualized images of women and girls returned for searches including "Black," "Latina," and

But what does this have to do with agency and information? After all, as Sweeney points out, the ads themselves may be well-attuned to their audiences, and it might be that search engines have a responsibility to ensure that their targeted advertising does not reflect race simply on the basis of harm prevention. But our argument here is different. It is that people have claims to information about some kinds of algorithmic systems even where their individual stake is relatively small, even where the system is reliable, and even where the system makes no assumptions about responsibility. So, while people who are White have relatively little *personal* stake in the issue of search engine advertising serving ads that suggest arrest records disproportionately to searches using Black-identifying names, they have an interest based in agency nonetheless. Specifically, they have an interest in exercising agency over areas of democratic concern.

For the moment we will call this *democratic agency*, and define it as access to information that is important for persons to perform the legitimating function that is necessary to underwrite democratic authority. We will take up this facet of agency and autonomy in more detail in chapter 8. The gist of the idea is this. Whether a democratic state, set of policies, actions, regulatory regimes, and so forth is justifiable (or *legitimate*) is in important part a function of the autonomy of its citizens. Exercising the autonomy necessary to serve this

"Asian," but not "White," searches for high-status positions returning images predominantly of White people (e.g., "CEO"), facial recognition and image enhancement technologies that are more accurate for images of White people than Black people, health risk assessment machine learning tools that underestimate Black patients' eligibility for care interventions, and more. Jonathan Garvie and Clare Frankle, "Facial-Recognition Software Might Have a Racial Bias Problem," *The Atlantic*, April 7, 2016, https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/; Safiya Umoja Noble, *Algorithms of Oppression : How Search Engines Reinforce Racism* (New York: New York University Press, 2018); Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (October 25, 2019): 447–53. While organizations often aim to rectify these disparities, those responses are often reactive. Moreover, knowledge of those processes is important to democratic agency and legitimation, the topic of chapter 8.

legitimating function requires certain kinds of information. Google of course is not a state actor, but it serves an outsized role in modern life, and understanding how that interacts with basic rights (including treatment of people based on race), is important for people to understand.

4.3. Relation to the GDPR

Having examined moral claims to information about algorithmic systems based on cognitive and practical agency, it will be useful to consider some of the scholarship on *legal* rights to information regarding algorithmic systems. Specifically, there is considerable scholarly discussion regarding informational rights in the context of the European Union's General Data Protection Regulation.[42] Much of that discussion concerns whether the GDPR contains a "right to explanation," and if so, what that right entails. There is, in contrast, much less scholarly attention devoted to what moral claims (if any) underwrite such a right. The claims to cognitive and practical agency that we have established can do that justificatory work. But before we get to that, we want to draw on some of the right to explanation scholarship for some important context and to make a few key distinctions.

The General Data Protection Regulation (GDPR) is the primary data protection and privacy regulation in European Union law. For our purposes, we wish to discuss four specific rights related to decision systems: *the right of access* (the right to access the information in one's file), *the right to rectification* (the right to correct misinformation in one's file), *the right to explanation* (the right to have automated decisions made about oneself explained), and *the right*

[42] European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)," 2016 O.J. (L 119) 1 § (2016).

to object (the right not to be subject to a significant decision based solely on automated processing).

4.3.1 The right of access and the right to rectification

Article 15 of the GDPR outlines the *right of access*, which is the (legal) right of data subjects who are citizens of the EU to obtain from data controllers confirmation as to whether or not their personal data are being processed, confirmation that personal data shared with third parties is safeguarded, and to obtain a copy of personal data undergoing processing.[43] Article 16 outlines the *right to rectification*, which is "the [legal] right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her."[44] These legal rights can be underwritten by the same ideas that support the principles of practical and cognitive agency and the principles of informational control, and we can use the principles to underwrite them.

Begin with rectification. Where one's data is being used to make decisions affecting significant facets of one's life—such that the system restricts one's agency—the principle of informational control tells us that there is a defeasible claim to correcting that information. Insofar as our data is being used to make decisions about us that will affect us, the right to rectification stands as a law that enjoys justification from this principle.

With these ideas in place, we can also offer a justification for the right of access. To know whether a controller has incorrect information about us or information that we do not want

[43] European Union, pt. 15.

[44] European Union, pt. 16.

them to have or share, we need to know what information they in fact have about us. And so, if the right to rectification is to have value, we need a right of access. We can further support the right of access by reflection of the principles of practical and cognitive agency: often we will need to know what information is being collected in order to improve our prospects or to simply make sense of decisions being made about us.

4.3.2 The right to explanation

Consider next the right to have significant automated decisions explained. The Arroyo case brings out the importance of this right. To respond to their predicament, Carmen and Mikhail need to understand it. We begin with a general discussion of the right.

Sandra Wachter, Brent Mittelstadt, and Luciano Floridi introduce two helpful distinctions for thinking about the right to explanation.[45] The first of these distinctions disambiguates *what* is being explained. A "system-functional" explanation explains "the logic, significance, envisaged consequences and general functionality of an automated decision-making system."[46] In contrast, a "specific decision explains "the rationale, reasons, and individual circumstances of a specific automated decision."[47] Note that if a system is deterministic a complete description of system functionality might entail an explanation of a specific decision. So, in at least some cases, the distinction between the two kinds of explanation is not exclusive.[48]

[45] Sandra Wachter, Brent Mittelstadt, and Luciano Floridi, "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation," *International Data Privacy Law* 7, no. 2 (May 1, 2017): 76–99.

[46] Wachter, Mittelstadt, and Floridi, 11.

[47] Wachter, Mittelstadt, and Floridi, 11.

[48] Andrew D. Selbst and Julia Powles, "Meaningful Information and the Right to Explanation," *International Data Privacy Law* 7, no. 4 (November 1, 2017): 233–42.

The second distinction disambiguates *when* the explanation is being given. An *ex ante* explanation occurs prior to when a decision has been made. An *ex post* explanation occurs *after* the decision has been made. Wachter et al. claim that ex ante explanations of specific decisions are not possible; a decision must be made before it is explained. As Andrew Selbst and Julia Powles point out, in the special case of a complete system-level explanation of a deterministic system, decisions are predictable and thus, ex ante explanations of those decisions are at least sometimes possible.[49]

Rather than stake a claim in this dispute, we will take a pragmatic approach. We can say all we need to say about the right to explanation by discussing the three categories that Wachter et al. admit of (i.e., ex ante system-functional, ex post system functional, and ex post specific). If a subject has a right to an ex ante explanation of a specific decision, the arguments for such explanations will follow naturally from our arguments for specific explanations; the only issue that the right will turn on is whether such explanations are possible—an issue that we are not taking a stand on here. We think that, morally, the right to explanation could encompass any of the possibilities Wachter et al. outline. So, we will understand the right to explanation as the right to explanations about ex post specific decisions, ex ante system function, or ex post system function.

Let us then work though some ideas about what our account says about the right to explanation.

[49] Selbst and Powles.

Ex ante system-functional explanations. Subjects of decisions that have not yet been made often have good reason to know how decisions of that sort will be made in the future. The principles of practical agency delineate some of these conditions.

One way to see this is to return to Catherine Taylor. She now knows that because of her common name, systems that perform quick, automated searches are prone to making mistakes about her. Based on this, she has an interest in knowing how a given system might produce a report on her. If she knows a system is one that might produce a false report about her, she can save herself—and the purchaser of the report—quite a bit of trouble, either by insisting to ChoicePoint that more careful methods are used, or by preempting the erroneous results by providing an independent, high quality counter-report of her own.

Ex post system-functional explanations: Subjects of decisions that have been made often have good reason to know how those decisions of that sort were made. These claims can be grounded in practical or cognitive agency.

Consider Taylor again. If Taylor is denied a job and she learns that an automated background check was involved, she has reason to suspect that the automated check might have erroneously cost her the opportunity. For her, simply knowing the most general contours of how a system works is powerful information. This alone may be enough to allow her to get her application reviewed again, and she could not reasonably endorse a system where she is denied this minimal amount of information. But even if she cannot accomplish this—that is, even if the principle of informed practical agency is not activated because her situation is hopeless—she still has a claim, via the principle of informed cognitive agency, to gain an understanding of her situation.

Specific explanations: Finally, subjects of decisions often have good reason to know how those specific decisions were made. These claims can be grounded in practical agency.

Recall Arroyo's denial of housing. Something is wrong with Arroyo's report, yet his mother does not (and cannot) know what it is. This leaves her especially vulnerable in defending her son, since she does not know what to defend him against. As the principle of informed practical agency demands, subjects of decisions that have been made should at least know enough about those decisions to respond to them if they have been made in error.

We want to pause briefly to discuss a recent proposal pertaining to *how* specific explanations might be given, namely, via *counterfactual explanations*, which have been detailed extensively in a recent article by Wachter et al. An example of a counterfactual explanation, applied to the Arroyo case, is as follows

> "You have been denied tenancy because you have one criminal charge in your history; Were you to have had zero charges, you would have been granted tenancy."

Generalizing a bit, counterfactual explanations are explanations of the form *"W* occurred because *X*; Were *Y* to have been the case (instead of *X*), *Z* would have occurred (instead of *W*)," where *W* and *Z* are decisions and *X* and *Y* are two "close" states of affairs, identifying a small—perhaps the smallest—change that would have yielded *Z* as opposed to *W*.

Counterfactual explanations have several virtues qua specific explanations. For one, they are easy to understand.[50] They are efficient in communicating the important information users need to know to make sense of and respond to decisions that bear on them. Thus, such explanations are often sufficient for giving subjects what they are informationally owed. Another virtue is that they are relatively easy to compute, and so producing them at scale isn't onerous: algorithms can be written for identifying the smallest change that would have made a difference with respect to the decision.[51] Further, they communicate needed information without compromising the algorithms that underlie the decisions they explain; they offer explanations, as Wachter et al put it, "without opening the black box."[52]

Counterfactual explanations can serve as a useful tool for delivering what is demanded by the cognitive and practical agency of data subjects without running roughshod over the interests of their data controllers. Of course, such explanations won't *always* meet these demands; they will only work in contexts where specific explanations are called for. And even then, they might not *always* offer everything an agent needs; for instance, one could imagine counterfactual explanations that are too theory laden to be useful, or one that is only informative against knowledge of myriad background conditions. Nevertheless, this style of explanation *can* be a very useful tool in meeting agents' needs. Thus, they serve as a good example of a realistic tool for giving data subjects what they are informationally owed.

[50] S. Wachter, B. D. M. Mittelstadt, and C. Russell, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR," *Harvard Journal of Law and Technology* 31, no. 2 (2018): 841–87.

[51] Wachter, Mittelstadt, and Russell, 15–16.

[52] Wachter, Mittelstadt, and Russell, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR."

Let’s take stock of what our account has to say about the right to an explanation. We take it that the right to explanation is a defeasible right to meaningful ex post, ex ante, system-level and specific explanations of significant, automated decisions. Using our cases and principles, we have demonstrated how our account can underwrite a claim: as autonomous beings, we need to understand significant events in our lives in order to navigate the world so as to pursue our values; as autonomous beings, we have a duty to support each other's autonomy; so, if we are in control of information pertaining to significant decisions affecting someone’s life, we often owe it to them to make that information available.

4.3.3 The right to object

In addition to rights of access, rectification, and explanation, the GDPR outlines the *right to object*, “the right not to be subject to a [significant] decision based solely on automated processing.”[53] As above, our interest is in understanding whether there is a moral right to object. However, examining a version of a legal right can help us make sense of moral claims. There are two key features of the right to object as it is stated in the GDPR.

Note first that the right is vague. Specifically, the “based solely” condition, as well as the notion of significance, admits of vagueness. As Kaminski notes,

> One could interpret “based solely” to mean that any human involvement, even rubber-stamping, takes an algorithmic decision out of Article 22’s scope; or one could take a broader reading to

[53] European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), pt. 22.

> cover all algorithmically-based decisions that occur without meaningful human involvement. Similarly, one could take a narrow reading of "[…] significant" effects to leave out, for example, behavioral advertising and price discrimination; or one could take a broader reading and include behavioral inferences and their use.[54]

We will not focus too heavily on issues of vagueness here. However, it is important to note that the limiting condition of the right—as well as some of its content—is vague.

Second, the right to object is ambiguous.[55] It could be understood broadly: as a broad prohibition on decisions that are based solely on automated processing. The same right could also be understood narrowly: as an individual right that data subjects can summon, for the purposes of rejecting a particular algorithmic decision.[56] Here, we won't be interested in adjudicating which way to read Article 22 of the GDPR, because we regard both readings as supported by the same considerations that we cite in favor of the right to explanation.

Human oversight of an automated decision system requires that the system be functionally intelligible to at least some humans (perhaps upon acquiring the relevant expertise). So, in a world where the broad reading is observed, each significant automated decision is intelligible to some human overseers. What this means, in turn, is that the reasons for its

[54] Margot E. Kaminski, "The Right to Explanation, Explained," *Berkeley Technology Law Journal* 34, no. 189 (2019).

[55] Kaminski; Isak Mendoza and Lee Bygrave, "The Right Not to Be Subject to Automated Decisions Based on Profiling," in *EU Internet Law: Regulation and Enforcement*, ed. Tatiani-Eleni Synodinou et al., 2017, 77–98.

[56] The terminology of broadness and narrowness is from Kaminski, "The Right to Explanation, Explained."

decisions could be meaningfully explained to data subjects (or at the very least to their surrogates). The significance of this, from our point of view, is that it would help to secure the right to explanation, as it would require systems to be designed so that they are intelligible to humans. Further, in cases where a data subject cannot request an explanation, it serves to ensure them that significant automated decisions made about them make sense. Similarly, in a world where the narrow right is observed, systems are designed to be intelligible so that, *were* their decisions meaningfully checked by a human decision maker, they *would* make sense. This, of course, means that they are designed so that they do make sense to humans (even if those humans are experts). Further, like the broad reading, it also affords data subjects the opportunity to have decisions checked when they themselves cannot check them (perhaps for reasons of trade secrecy). However, the narrow right might sound more plausible than the broad right because it means fewer human decision makers would have to be employed to satisfy it, allowing systems to operate more efficiently.

Now, unlike the previously mentioned rights, the right to object—particularly in its broad formulation—might sound onerous. However, abiding the rights to access, rectification, and explanation already requires that data controllers provide data subjects meaningful human oversight of decision made about them, so perhaps the broad right isn't as implausible as it may first seem. Further, the broad right has the advantage that it makes the exercise of the right to object less costly to those individuals who would otherwise have to explicitly exercise it. We can imagine data subjects worrying that they will face prejudice for exercising the right; for instance, a job applicant might worry that if she exercised the right, the potential employer will think that she is going to cause trouble.

What does the right to object add, then? Importantly, there are systems where inferences must be kept secret—either to prevent subjects from gaming it, or because the system is simply too complicated—in these circumstances, the right to object plays the important role of ensuring that surrogates of data subjects understand whether high-stakes decisions made about those subjects make sense.

4.4. Polestar cases

We can finally return to the cases that provide our through-line through the book.[57]

4.4.1. *Loomis*

One of Loomis's primary complaints in his appeal is that COMPAS is proprietary and hence not transparent. Specifically, he argued that this violated his right to have his sentence based on accurate information. He bases the argument in part on *Gardner v. Florida.*[58] In *Gardner*, a trial court failed to disclose a presentence investigation report that formed part of the basis for a death sentence. The U.S. Supreme Court determined that the failure to disclose the report meant that there was key information underwriting the sentence which the defendant "had no opportunity to deny or explain." Loomis argued that the same is true of the report in his case. Because the COMPAS assessment is proprietary[59] and because there had not been a validation study of COMPAS's accuracy in the state of Wisconsin (other states had conducted validation studies of

[57] For fuller description of our polestar cases see chapter 1 and see Alan Rubel, Clinton Castro, and Adam Pham "Algorithms, Agency, and Respect for Persons," *Social Theory & Practice* 46(3) (July 2020): 547-572; https://doi.org/10.5840/soctheorpract202062497.

[58] Gardner v. Florida, 430 U.S. 349 (1977).

[59] Wisconsin v. Loomis, 881 N.W.2d 749 (Wisconsin Supreme Court 2016).

the same system), Loomis argued that he was denied the opportunity to refute or explain his results.

The Wisconsin supreme court disagreed. It noted that Northpointe's Practitioner's Guide to COMPAS explained the information used to generate scores, and that most of the information is either static (e.g., criminal history) or in Loomis's control (e.g., questionnaire responses). Hence, the court reasoned, Loomis had sufficient information and the ability to assess the information forming the basis for the report, despite the algorithm itself being proprietary.[60] As for Loomis's arguments that COMPAS was not validated in Wisconsin and that other studies criticize similar assessment tools, the court reasoned that cautionary notice was sufficient. Rather than prohibiting use of COMPAS outright, the court determined that presentence investigation reports using COMPAS should include some warnings about its limitations.

According to the principles of practical agency, Loomis has a defeasible claim to information about COMPAS if (a) information about COMPAS advances his practical agency, (b) because COMPAS has restricted his practical agency, (c) COMPAS's effects bear heavily on significant aspects of Loomis's life, and (d) information about COMPAS allows Loomis to correct or mitigate the effects of COMPAS. If there is such a claim, it is strengthened (e) if COMPAS purports to be based on factors for which Loomis is responsible and (f) if COMPAS has errors.

It is certainly plausible that COMPAS limits Loomis's practical agency insofar as it had some role in his sentence. Loomis faced a number of decisions about what to do in response to

[60] Wisconsin v. Loomis, 881 N.W.2d paragraphs 54–56.

his sentence. One is whether he should appeal and on what grounds. Another is whether he should try to generate public support for curtailing use of COMPAS. For Loomis, settling these questions about what to do depends on knowing how COMPAS generated his risk score. And there is much he doesn't know. He doesn't know whether the information fed into COMPAS was accurate. He doesn't know whether, and in what sense, COMPAS is fair. And he doesn’t know whether the algorithm was properly applied to his case. That lack of information curtails his practical agency. The length of his criminal sentence certainly involved a significant facet of his life, and it is at least plausible that greater information would allow him to mitigate COMPAS’s effects. The strength of his claims increases in light of the fact that it is best understood as being based on factors for which he is responsible, viz., his propensity to re-offend.

So, Loomis has a prima facie and defeasible claim to information about COMPAS. But that leaves open just what kind of information he has a claim to, what that claim entails, and whether there are countervailing considerations that supersede Loomis’s claim. It would seem that Loomis needs to know that the data fed into COMPAS was accurate, evidence that COMPAS is in fact valid for his case, and, finally, some kind of explanation—perhaps in the form of a counterfactual explanation—that makes clear why he received the score that he did. Such information would advance Loomis’ practical agency, either by giving him the information needed to put together an appeal *or* by demonstrating to his satisfaction that his COMPAS score was valid, allowing him to focus his efforts elsewhere.

Independent of the concerns based on practical agency, Loomis has a claim to information based on cognitive agency. Both factors in the Principle of Informed Cognitive Agency are present. COMPAS purports to be based on factors for which Loomis is responsible, and the stakes are high. Being imprisoned is among the most momentous things that may happen to a person and understanding the basis of a prison sentence is essential to one's agency. That extends beyond the factors that matter in determining one's sentence to include whether the process by which one is sentenced is fair. And as we have argued, agents have a claim to understand important facets of their situations. Hence, Loomis has a claim based on cognitive agency to better understand the grounds for his imprisonment.

While Loomis plausibly has claims to information based on both practical and cognitive agency, there are differences in what those claims entail. While practical agency will only underwrite information that can be used in advancing Loomis's case—and hence, mostly supports information for Loomis's legal representation—cognitive agency underwrites the provision of certain pieces of information to Loomis himself. It would involve providing him information about the fact that a proprietary algorithm is involved in the system, information about how well the system predicts re-offense, and information about the specific factors that led to Loomis's sentence. There is no reason to think that it would advance Loomis's cognitive agency to provide him with specific information about how COMPAS functions.

Moreover, the court did, in fact, respect Loomis's cognitive agency. The Wisconsin Supreme Court upheld the circuit court's decision in substantial part because the circuit court articulated its own reasons for sentencing Loomis as it did. In other words, it provided an account sufficient for Loomis to exercise evaluative control with respect to his reactive attitudes toward the decision and sentence.

4.4.2. *Wagner* and *Houston Schools*

The principles of informed practical agency and informed cognitive agency also aid our understanding of the K-12 teacher cases, especially *Houston Schools*. Recall that Houston Schools uses a VAM called EVAAS, which produces each individual teacher's score by referencing data about all teachers.[61] This practice makes EVAAS's scores highly interdependent. Recall also that Houston Schools was frank in admitting that it would not change faulty information because it would require a costly re-analysis for the whole school district, and the potential to change all teachers' scores. This all was despite warnings (as we note in chapters 1 and 3) that value added models have substantial standard errors.[62] So, EVAAS's scores are extremely fragile, produced without independent oversight, and cannot be corroborated by teachers (or the district or, recall, an expert who was unable to replicate them).

It seems clear enough that information about EVAAS is vital for teachers to exercise practical agency. Certainly, it is relevant to several significant aspects of teachers' lives. For teachers who were fired or did not have their contracts renewed based on low performance, gaining an understanding of the system advances their practical agency in a couple of ways. It gives them (and their union leaders and lawyers) the bases of either an appeal (whether in court or to the public) of the firings or an appeal of the system altogether. It also gives teachers who are finding employment in other schools some context that could help them convince administrators that their departure from HISD was not evidence of poor teaching. That is, affected teachers have a (defeasible) claim to information about EVAAS's functioning, because

[61] Houston Fed of Teachers, Local 2415 v. Houston Ind Sch Dist, 251 F. Supp. 3d 1168 (S.D. Tex. 2017).

[62] David Morganstein and Ron Wasserstein, "ASA Statement on Value-Added Models," *Statistics and Public Policy* 1, no. 1 (December 22, 2014): 7.

it could allow them to correct or mitigate the system's effects. Their claim is strengthened because EVAAS purports to be based on factors for which the teachers are responsible (viz., their work in the classroom), and yet (as HISD admits) EVAAS has errors. These claims also underwrite teachers' claims to informational control, specifically their claim to have any inaccuracies reflected in their scores corrected.

The fact that EVAAS affects such important parts of teachers' lives and purports to be based on factors for which they are responsible also gives them a claim to information based in cognitive agency. As in the COMPAS case, the *type* of information necessary for teachers to exercise evaluative control—that is, to assess their treatment at the hands of their school system—may be different from the information necessary for them to exercise practical agency. Cognitive agency may only require higher-level information about how EVAAS works, a frank assessment of its flaws, and a candid accounting of Houston Schools' unwillingness to incur the cost of correcting errors rather than the more detailed information necessary for teachers to correct errors. To put a bookend on the importance of cognitive agency, we will return to an exemplary teacher's public reaction to the VAM used by DC Schools: "I am baffled how I teach every day with talent, commitment, and vigor to surpass the standards set for me, yet this is not reflected in my final IMPACT score."[63] This would seem to be an appeal to exercise evaluative control.

[63] Valerie Strauss, "D.C. Teacher Tells Chancellor Why IMPACT Evaluation Is Unfair," *Washington Post*, August 16, 2011, sec. Local.

4.5. Conclusion

In chapter 2, we argued that autonomy ranges beyond the ability to make choices. Properly understood, self-governance includes competence and authenticity and substantive independence, and it demands acting accord with others. Chapter 3 examined the requirements for respecting persons' autonomy related to their material conditions. In the present chapter, we explain the informational requirements of autonomy. Specifically, we argued that autonomy demands respect for both practical and cognitive agency. We articulated several principles of practical and cognitive agency and argued that those principles could underwrite key provisions in the GDPR. Finally, we explained that those principles entail that the subjects of our polestar cases deserve substantial information regarding the algorithmic systems to which they are subject.

Recall, though, that the organizing thesis of the book is that understanding the moral salience of algorithmic systems requires understanding how such systems relate to autonomy. That involves more than respecting the autonomy of persons who are, at the moment, autonomous. It also involves securing the conditions under which they can actually exercise autonomy. That's the issue we turn to in next two chapters.

References

91st United States Congress. An Act to amend the Federal Deposit Insurance Act to require insured banks to maintain certain records, to require that certain transactions in U.S.

currency be reported to the Department of the Treasury, and for other purposes., 15 U.S.C. § 1681 § (1970).

Connecticut Fair Hous. Ctr. v. Corelogic Rental Prop. Sols., LLC, 369 F. Supp. 3d 362 (D. Conn. 2020).

European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1 § (2016).

Gardner v. Florida, 430 U.S. 349 (1977).

Garvie, Jonathan, and Clare Frankle. "Facial-Recognition Software Might Have a Racial Bias Problem." *The Atlantic*, April 7, 2016. https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/.

Hill, Jr., Thomas. "Autonomy and Benevolent Lies." *The Journal of Value Inquiry* 18, no. 4 (December 1, 1984): 251–67.

Houston Fed of Teachers, Local 2415 v. Houston Ind Sch Dist, 251 F. Supp. 3d 1168 (S.D. Tex. 2017).

Kaminski, Margot E. "The Right to Explanation, Explained." *Berkeley Technology Law Journal* 34, no. 189 (2019).

Kirchner, Lauren, and Matthew Goldstein. "Access Denied: Faulty Automated Background Checks Freeze Out Renters," May 28, 2020. https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters.

Mendoza, Isak, and Lee Bygrave. "The Right Not to Be Subject to Automated Decisions Based on Profiling." In *EU Internet Law: Regulation and Enforcement*, edited by Tatiani-Eleni Synodinou, Philippe Jougleux, Christiana Markou, and Thalia Prastitou, 77–98, 2017.

Morganstein, David, and Ron Wasserstein. "ASA Statement on Value-Added Models." *Statistics and Public Policy* 1, no. 1 (December 22, 2014): 108–10.

Nelson, Ariel. "Broken Records Redux: How Errors by Criminal Background Check Companies Continue to Harm Consumers Seeking Jobs and Housing." Boston, MA: National Consumer Law Center, December 6, 2019.

Noble, Safiya Umoja. *Algorithms of Oppression : How Search Engines Reinforce Racism*. New York: New York University Press, 2018.

Nyholm, Sven. *Humans and Robots: Ethics, Agency, and Anthropomorphism*. London; New York: Rowman & Littlefield Publishers, 2020.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366, no. 6464 (October 25, 2019): 447–53.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. 1st ed. New York: Crown, 2016.

Oshana, Marina. *Personal Autonomy in Society*. 1st ed. Aldershot, Hants, England ; Burlington, VT: Routledge, 2006.

Rubel, Alan. "Privacy and the USA Patriot Act: Rights, the Value of Rights, and Autonomy." *Law and Philosophy* 26, no. 2 (2007): 119–159.

Rubel, Alan, Clinton Castro, and Adam Pham. "Algorithms, Agency, and Respect for Persons." *Social Theory & Practice* 43, no. 3 (July 2020): 547–72.

Selbst, Andrew D., and Julia Powles. "Meaningful Information and the Right to Explanation." *International Data Privacy Law* 7, no. 4 (November 1, 2017): 233–42.

Smith, Michael. "A Constitutivist Theory of Reasons: Its Promise and Parts." *Law, Ethics and Philosophy* 1 (December 1, 2013): 9–30.

Solove, Daniel J. "Privacy Self-Management and the Consent Dilemma." *Harvard Law Review* 126 (2013): 1880–1903.

Strauss, Valerie. "D.C. Teacher Tells Chancellor Why IMPACT Evaluation Is Unfair." *Washington Post*, August 16, 2011, sec. Local.

Sweeney, Latanya. "Discrimination in Online Ad Delivery." *Communications of the ACM* 56, no. 5 (January 28, 2013): 44–54.

Wachter, S., B. D. M. Mittelstadt, and C. Russell. "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR." *Harvard Journal of Law and Technology* 31, no. 2 (2018): 841–87.

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation." *International Data Privacy Law* 7, no. 2 (May 1, 2017): 76–99.

Whittlestone, Jess, Rune Nyrup, Anna Alexandrova, Kanta Dihal, and Stephen Cave. "Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: A Roadmap for Research." London: The Nuffield Foundation, 2019.

Wisconsin v. Loomis, 881 N.W.2d 749 (Wisconsin Supreme Court 2016).

Yu, Persis S., and Sharon M. Dietrich. "Broken Records: How Errors by Criminal Background Checking Companies Harm Workers and Businesses." National Consumer Law Center, April 11, 2012.