Algorithms, Agency, and Respect for Persons

Alan Rubel (arubel@wisc.edu), University of Wisconsin-Madison

Clinton Castro (clinton.g.m.castro@gmail.com), Florida International University

Adam Pham (adamkpham@gmail.com), University of Wisconsin-Madison

Abstract

Algorithmic systems play an increasingly important role in modern life. Much of the scholarship on the moral ramifications of such systems focuses on bias and harm. We argue that understanding the moral salience of algorithmic systems requires understanding the relation between algorithms, autonomy, and agency. We use recent cases in criminal sentencing and K-12 teacher evaluation to outline four key ways in which issues of agency, autonomy, and respect for persons can conflict with algorithmic decision-making. Three involve failures to treat individuals with sufficient respect. The fourth involves distancing oneself from morally suspect actions by laundering one's agency.

Keywords

algorithm; agency; autonomy; big data; agency laundering; COMPAS

1. Introduction*

In *Wisconsin v. Loomis*, Eric Loomis pleaded guilty to crimes related to a drive-by shooting.[1] The trial judge ordered a presentence investigation report , using a proprietary, algorithmic risk

* The authors would like to thank a number of people who have commented on drafts or presentations related to this project, including Richard Warner, Robert Streiffer, Filippo Santoni de Sio, Sven Nyholm, Tijn Borghuis, Suresh Venkatasubramanian, Owen King, Jeroen van den Hoven, Marc Alfano, the BIAS workshop at University of Sheffield, the 4TU Centre for Ethics and Technology (Netherlands), the departments of philosophy at Eindhoven University of Technology, University of Twente, Delft University of Technology, and audiences at the Amsterdam Privacy Law Scholars Conference, the University of Texas School of Information, and the 2018 Information Ethics Roundtable. We also want to thank the anonymous Reviewer 1 for *Social Theory & Practice*, whose comments were especially helpful and professional.

The authors have more thoroughly articulated the conception of agency laundering (section 6) in a separate article (Rubel, Castro, and Pham (2019a)) and conference proceedings (Rubel, Castro, and Pham (2019b). The concept is explained here in service of a broader point about the relationship between algorithms, agency, and autonomy rather than a fully-developed and applied account.

[1] 2016 WI 68, 881 N.W.2d 749, 757, 760-64, cert. denied, 582 U.S. _ (U.S. June 26, 2017)(No. 16-6387).

Pre-production copy. Please cite to final version: *Social Theory and Practice* 46(3)(2020): xx-xx, https://doi.org/10.5840/soctheorpract202062497.

assessment tool called COMPAS that was developed by Northpointe, Inc.[2] COMPAS is not designed for determining offender sentences, and Northpointe specifically warns against using it in sentencing decisions. Nonetheless, the judge used the presentence investigation report and COMPAS assessment while sentencing Loomis in the maximum range.

Much of the growing literature on algorithms focuses on harm, discrimination, and inscrutability; but *Loomis* presents a puzzle. It is plausible that Loomis was not harmed in that he received exactly the sentence he would have received without the presentence investigation report. Moreover, because he is white, and the algorithm appears to disadvantage black defendants, the judge's use of the report did not expose Loomis to racial discrimination. Nonetheless it seems that he was wronged. But how so?

We will argue (inter alia) that Loomis should be able to understand the mechanisms by which he is evaluated for incarceration and to use that understanding to make his case. Denying him that understanding is not a harm in itself (though it may result in a harm), but a failure of respect for him as a person. One might argue that the primary issue here is the need for transparency in the algorithm. However, absent an explanation for why transparency matters, the argument is incomplete.

Now consider a different, seemingly unrelated case. In 2007, Washington, D.C., sought to improve its public-school system (“DC schools”) by implementing an algorithmic teacher-

[2] COMPAS (“Correctional Offender Management Profiling for Alternative Sanctions") is a suite of assessment tools developed for use at various stages in the criminal justice system with different algorithms and software packages focusing on (inter alia) defendants who are recently incarcerated or under supervision (COMPAS Core), persons who will soon reenter their community after incarceration (COMPAS Reentry), young people (COMPAS Youth), and general case management (Northpointe Suite Case Manager). The tool used in Loomis is COMPAS Core (which we call "COMPAS" for simplicity). Northpointe is now a part of Equivant, Inc. Equivant, Inc., “Northpointe Suite Automated Decision Support,” available at http://www.equivant.com/assets/img/content/Northpointe_Suite_2.pdf.

assessment tool, IMPACT,[3] that aimed to identify and eliminate ineffective teachers. In 2010, teachers with IMPACT scores in approximately the bottom 2% were fired; the following year, teachers with IMPACT scores in approximately the bottom 5% were fired (O'Neil 2016; Turque 2011). However, IMPACT has epistemic flaws. These flaws are serious enough that the American Statistical Association has warned that rankings created by such tools are “unstable” and overwhelmingly attributable to factors other than teaching (ASA 2014).

The problem with IMPACT (if there is one) is not simply that it harms teachers. Firing teachers can be justified, though it harms them. And IMPACT is not obviously unfair; its epistemic flaws may be evenly distributed among teachers, and may create enough benefit for students to risk some harm to teachers. Nonetheless, *something* seems off about using IMPACT.

Our contention is that many algorithmic systems[4] are similar to the ones in *Loomis* and IMPACT in that they engender wrongs that are not best understood as harms.[5] More specifically, we argue that a complete picture of the moral salience of algorithmic systems requires understanding algorithms as they relate to agency, autonomy, and respect for persons. In other words, quite different algorithmic systems, which are used in quite different settings, fail to respect persons’ autonomy in similar ways.

[3] Perhaps surprisingly, IMPACT is not an acronym.

[4] The term "algorithm" can be used to describe any abstract decision-making procedure. In practice, however, the term typically refers to a systematic implementation (Mittelstadt et al. 2015, 2) of an algorithm, or an "algorithmic decision system," such as COMPAS.

[5] Put another way, these wrongs are not reducible to harms to material interests, physical or psychological damage, or discrimination (though algorithmic systems can cause such harms). The wrongs could be considered harms, however, if we understand ‘harm’ in the broadest sense such that it includes harms to one’s moral interests as a human being. Thanks to an anonymous reviewer for making this point.

Questions surrounding data-driven, automated, and inscrutable decision systems are vitally important. After the Wisconsin Supreme Court ruled in the *Loomis* case, the U.S. Supreme Court denied *certiorari*. Along the way, the case received national and international attention. *Loomis* is among the first cases in which courts have examined how algorithmic decision-making affects persons' legal rights. Likewise, a number of cases have also been brought challenging the use of proprietary, algorithmic evaluations of K-12 teachers. And there will doubtless be others. The arguments we advance here help understand the underlying moral issues involved in those cases.[6]

The paper is organized as follows. In section 2, we offer an account of agency and autonomy. And in sections 3-6 we draw on those examples to explain distinct ways that algorithmic systems conflict with agency and respect for persons, and we apply those explanations to a variety of recent legal cases contesting the use of algorithmic decision making. Specifically, we argue that algorithmic systems may govern behavior in ways that agents cannot reasonably endorse (sec. 3), may deny agents information to which they are entitled (sec. 4), may fail to respect boundaries between persons (sec. 5), and may allow agents to "launder" their morally suspect actions by attributing the decisions to the algorithms (sec. 6). We conclude in section 7 with some caveats and directions for further work.

[6] To be clear, we are offering a moral argument regarding the cases, not an analysis of legal doctrine. Using legal disputes here does several things. First, it demonstrates that the issues we are considering are vitally important public issues. Second, the cases offer rich, real-world examples to help understand the moral issues we are considering. And, lastly, they demonstrate that in the absence of better policy making and policy arguments, legal disputes may lead to morally problematic outcomes, even if those outcomes are consistent with existing law.

2. Agency and Autonomy

There is a growing literature concerning predictive analytics and algorithmic decision systems. Much of the literature focuses on how such systems should be treated legally (Citron 2008, Zarsky 2016, Barocas and Selbst 2016). Others have argued that such systems are too unreliable to deploy in meaningful settings (O'Neil 2016, Warner and Sloan 2018). Still others have asserted that use of algorithmic decisions is unfair (Pasquale 2015, Eubanks 2018) or oppressive (Noble 2018). Several authors have considered whether algorithmic systems affect the degree to which decisions made by agents are autonomous or manipulated (Yeung 2017, Lanzing 2019; Susser, Roessler, and Nissenbaum 2019). Our view is that the relationship between algorithmic decision systems and autonomy is not primarily about effects on individual decisions. Rather, it is a more general issue of respect for persons and the responsibilities of agents.

Autonomy is a complex and contested concept, and a fully articulated view is neither possible nor appropriate here. However, we can offer a basic account that is substantial enough to support our arguments. Autonomy includes at least the ability to self-govern. The ability to self-govern includes the ability to develop one's own conception of value and sense of what matters, to   the values that will guide one's actions and decisions, and to make important decisions about one's life according to those values where one sees fit (Hill 1989, Brock 1987). This way of conceptualizing autonomy owes a great deal to Kant. The third ("autonomy") formulation of the Categorical Imperative is "the idea of the will of every rational being as a will giving universal law," and requires a person act according to a maxim of one's will and such that their will "could at the same time have as its object itself as giving universal law" (Kant, Groundwork of the Metaphysics of Morals, 4:432). Maxims are compelling, and one acts

Pre-production copy. Please cite to final version: *Social Theory and Practice* 46(3)(2020): xx-xx, https://doi.org/10.5840/soctheorpract202062497.

autonomously in following universal maxims, in that one both sets and follows them. In contrast to Kant’s conception of autonomy, however, it is crucial to note that the capacity to self-govern, the values an agent develops, and the ways in which she incorporates those values into her life are socially situated (Mackenzie and Stoljar 2000, 4). Developing one’s sense of what is important depends on social conditions that nurture the ability to do so (Oshana 2006, 90). Social structures may delimit the conceptions of value that are available for a person to draw upon in developing her sense of value. Persons’ abilities to incorporate their values into their important decisions will depend on what opportunities exist in the broader social context (Raz, 1986; Mackenzie 2008). The fact that self-governance is socially situated, however, does not undermine the importance of autonomy and agency. Rather, failures to nurture persons’ abilities to develop their agency and substantial constraints on options available for incorporating values into persons’ lives are moral problems in part *because* of the importance of autonomy (See, e.g., Superson 2005, Meyers 1987).

The capacity to self-govern—to develop and endorse one’s sense of value and act according to those values as one sees fit—grounds a number of moral claims. For example, deception is wrong (when it is) in part because it circumvents an agent's ability to make decisions according to her own reasons. And paternalism is an affront to autonomy because it supplants a person’s agency—her ability to act—based on a degree of distrust of that agency (Shiffrin 2000). Of course autonomy can underwrite moral claims only to the extent that it is used to ends that are compatible with others’ reasonable interests. Joel Feinberg, for example, distinguishes capacity and successful exercise of autonomy from “autonomy as ideal.” Autonomy as ideal recognizes that people can exercise autonomy badly (and hence facets of autonomy are not necessarily virtues) and that people are parts of larger communities. Hence, Feinberg explains, the ideal of an autonomous person requires that their self-governance be consistent with the autonomy of others in their community (Feinberg 1989, pp. 44-45). That reflects Kant's understanding that morally right action requires that action can coexist with everyone else's ability to exercise freedom under universal moral law (Kant, Metaphysics of Morals 6:230).

While there are myriad disputes about the concept, scope, and moral value of autonomy, the conception offered here is both minimal and compatible with a wide range of views. We will draw on it to explain other moral claims grounded in autonomy and implicated by algorithmic decision systems.

3. Procedures No Agent Could Reasonably Endorse

To understand our first argument, let's return to the IMPACT case. There is a plausible argument for DC schools using IMPACT. The algorithm uses complex, data-driven methods to find and eliminate inefficiencies, and it purports to do this in an objective manner. Its inputs are measurements of performance and its outputs are a strict function of those measurements. Whether a teacher has, say, ingratiated herself to administrators would carry little weight in the decision as to whether to fire her. Rather, it is (ostensibly) her effectiveness as a teacher that grounds that decision.

Nonetheless, DC schools' use of IMPACT was problematic. This is in part because IMPACT's conclusions were epistemically flawed. A large portion of a teacher's score is based on an Individual Value-Added measurement (IVA) of student achievement (Walsh and Dotter 2014), which seeks to isolate and quantify a teacher's individual contribution to student achievement on the basis of annual standardized tests (Isenberg and Hock 2012). IVAs are poorly suited for this task. DC teachers work in schools with a high proportion of low-income students. At the time IMPACT was implemented, even in the wealthiest of the city’s 8 wards (Ward 3) nearly a quarter of students were low-income, and in the poorest ward (Ward 8) 88% were low income (Quick 2015). As a recent article on IMPACT notes, low-income students face a number of challenges that influence their ability to learn:

> These schools’ student bodies are full of kids dealing with the toxic stress of poverty, leaving many of them homeless, hungry, or sick due to limited access to quality healthcare. The students are more likely to have an incarcerated parent, to be deprived of fresh or healthy food, to have spotty or no internet access in their homes, or to live in housing where it is nearly impossible to find a quiet place to study (Quick 2015).

Given the challenges of their students, it is not surprising that fewer teachers in Ward 8 than Ward 3 are identified by IMPACT as "high performing" (*ibid*.).

The effects of poverty are confounding variables that affect student performance on standardized tests. For this reason, we cannot expect IVAs—which use only annual test scores to assess a teacher's individual contribution to student achievement—to reliably find the signal of bad teaching through the noise of student poverty (cf. O'Neil 2016). Indeed, the American Statistical Association warns that studies on IVAs "find that teachers account for about 1% to 14% of the variability in test scores, and that the majority of opportunities for quality improvement are found in the system-level conditions" (ASA, 2014). The American Statistical Society also notes that "[IVAs] have large standard errors, even when calculated using several years of data. These large standard errors make rankings [of teachers] unstable, even under the best scenarios for modeling" (ASA, 2014).

So, IMPACT suffers from an *epistemic* shortcoming. Is there also a *moral* problem? And, if so, what does that have to do with the epistemic problem, and what does it have to do with agency? Based on our conception of autonomy, our argument is that a teacher who is fired is *wronged* when that firing is based on a system that she could not reasonably endorse.

As agents, teachers have a claim to incorporate their values into their lives as they see fit. And *respecting* them requires recognizing them as value-determiners, neither thwarting nor circumventing their ability to act according to those values without good reason. Moreover, as agents they are capable of abiding fair terms of social agreement (so long as others do too), and hence 'good reasons' will be reasons that they can abide as fair terms of cooperation (Scanlon 2000, Rawls 1999). Teachers can endorse those reasons as either consistent with their own values or endorse them as a manifestation of fair social agreement.

Now, what is it to thwart an agent's ability to act according to her values? One example, discussed above, is deceit where one precludes an agent's ability to understand circumstances relevant to her actions. Another way to thwart agency is to create conditions in which agents are not treated according to reasons that they could rationally endorse, were they to given the opportunity to choose how to be treated.[7] That is, precluding persons from acting according to their values (e.g., by deceit) or placing them in circumstances that they cannot endorse as fair is a failure of recognition of them as value-determiners, and a form of disrespect.

[7] As these examples show, there is some room for different interpretations of what count as "reasons that agents could endorse." Derek Parfit considers this issue at length in his discussion of the "consent principle" in *On What Matters* (2011, see especially chapter 8, section 24). Notice, though, that our argument here considers necessary conditions for such reasons, not sufficient reasons. Hence, we leave aside whether those conditions require the agents to be fully rational, aware of all relevant reasons, aware of all relevant facts, and so forth. Thanks to an anonymous reviewer for raising this issue.

IMPACT fails to respect teachers in exactly this way, for three inter-related reasons.[8]

Reliability. For reasons we have noted, IMPACT is an unreliable tool for the evaluation of teacher efficacy. Teachers, like any professionals, can reasonably endorse a system in which they are evaluated based on their efficacy. Through their training and professionalization, they have endorsed the value of educating students. Moreover, fair terms of social cooperation would require that truly ineffective teachers be identified in order to improve education. But because IMPACT is unreliable, there is reason to think that it misidentifies teachers as ineffective. Hence, teachers will be loath to endorse being evaluated by IMPACT.

Responsibility. IMPACT's lack of reliability is not the only way it fails to respect agency. Imagine a case where a teacher evaluation system reliably measures student learning. Two teachers score poorly in this year's assessment. One scores poorly because she did not assign curriculum-appropriate activities, while the other scores poorly because her classroom lacks air-conditioning. Only the first teacher is responsible for her poor scores. The second teacher's scores are based on factors for which she is not responsible. Teachers could not reasonably endorse such a system.[9]

Given the population many DC teachers were working with—underserved students—IMPACT cannot be understood as tracking only factors for which teachers are responsible. The

[8] To be clear, we think that violations across any of these three dimensions make use of an algorithm morally problematic. However, we do not think that these three dimensions are exhaustive; this list is not meant to be comprehensive. There may be other considerations, such as consideration of desert or fairness, that can play an important role in assessing the appropriateness of the use of an algorithm.

[9] Notice that in this example responsibility and reliability are both relevant. Teachers could reasonably endorse a system in which their jobs depend on factors for which they are not responsible—population decline, e.g. However, firing teachers whose scores suffer because of exogenous factors (lack of air conditioning) involves criteria that are not teachers' responsibilities and which are unreliable in making teaching better.

effects of poverty, abuse, bullying, illness, undiagnosed learning disabilities, and so on plausibly undermine teacher efficacy. Yet, teachers bear no responsibility for those impediments. So, even if the IVAs were reliable, teachers could not reasonably endorse their implementation.

Stakes. Perhaps the most important factor in determining whether agents can reasonably endorse an algorithmic decision system is the stakes involved. Suppose that an IVA system is set up to provide teachers with lots of information about their own practices but is not used for comparative assessment. The scores are shared with teachers privately and are not used for promotion and firing. Such a system might not be very reliable, or it might measure factors for which teachers are not responsible. Nonetheless, teachers might endorse it despite its limitations because the stakes are low. But if the stakes are higher (work assignments, bonuses, promotions), it's reasonable for the employees to want the system to track factors which can be reliably measured and for which they are responsible.

DC schools' use of IMPACT is high-stakes. Teachers rely on their teaching for a paycheck, and many take pride in what they do. They have sought substantial training and often regard educating students as key to their identities. Having a low IMPACT score might cost a teacher her job and career, and it may well undermine her self-worth. By agreeing to work in particular settings they have formed reasonable expectations that they can continue to incorporate those values into their lives, subject to fair terms of cooperation (e.g., that they do their work responsibly and well, that demand for their services continues, that funding remains available, etc.).

IMPACT does poorly on our analysis. It is not reliable, it evaluates teachers based on factors for which they are not responsible, and it is used for high-stakes decisions. These points are reflected in teacher reactions to IMPACT. For example, Alyson Perschke—a fourth-grade teacher in DC schools—alleged in a letter to Chancellor Kaya Henderson that IVA's are "unreliable and insubstantial" (Strauss 2011). Perschke did so well in her in-class observations that her administrators and evaluators asked if she could be videotaped as "an exemplar" (Strauss 2011). Yet, the same year her IVA dragged her otherwise flawless overall evaluation down to average. Remarking on this, she says "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" (Strauss 2011).

3.1 Legal Disputes

Use of algorithmic decision systems to evaluate teachers is not unique to DC schools, and our framework helps make sense of the moral issues underlying other cases.

Wagner v. Haslam. In 2010, the state of Tennessee began requiring that school systems evaluate teachers on the basis of IVAs. One IVA endorsed by the state legislature was the Tennessee Value-Added Assessment System (TVAAS), a proprietary system similar to IMPACT.

The TVAAS system includes standardized tests for students in a variety of subjects, including Algebra, English, Biology, Chemistry, and U.S. History. Roughly half of teachers at the time of the case taught in subjects not tested under TVAAS. Nonetheless, because of the law requiring teacher evaluation on the basis of IVAs, teachers of non-tested subjects were evaluated on the basis of a "school-wide composite score," which is the average performance of *all* students on *all* subjects in that school. In other words, it is a score that is identical for all teachers in the school regardless of what subjects and which students they teach.

Teresa Wagner and Jennifer Braeuner teach non-tested subjects (physical education and art, respectively). From 2010-2013, each received excellent evaluation scores based on observations of their individual classes combined with their schools' composite scores. In the 2013-14 school year, however, their schools’ composite scores dropped from the best possible score to the worst possible score, while their individual classroom observation scores remained excellent. The result was that Wagner's and Braeuner's individual, overall evaluations decreased from the highest possible to middling. This was enough to preclude Wagner from receiving the performance bonus she had received in previous years and to make Braeuner ineligible for consideration for tenure. Moreover, each "suffered harm to her professional reputation, and experienced diminished morale and emotional distress." (112 F. Supp. 3d at 690).

Wagner and Braeuner argued that the use of TVAAS and school-wide composite scores violated their due process and equal protection rights under the Fourteenth Amendment.[10] The court rejected those claims on the ground that the evaluation system met the rational basis standard.[11]

There is a deeper moral issue grounding the legal case. Wagner and Braeuner frame their case in terms of harms (losing a bonus, precluding tenure consideration, and so forth), but those harms matter only because they are wrongful. They are wrongful because TVAAS is an evaluation system that teachers could not reasonably endorse. Wagner and Braeuner's scores

[10] TVAAS's use of non-composite scores has also been challenged. See *Trout v. Knox County Board of Education* (163 F. Supp. 3d 492, E.D. Tenn. 2016) for details. Our framework also applies to *Trout*, but we do not have space to apply the framework.

[11] It is worth noting that rational basis is a very low standard. As the Court states "the review is limited to determining only whether there is a *conceivable* rational *relationship* between the policy and a legitimate governmental objective" 112 F. Supp. 3d at 692-93.

did not reliably track their performances nor did the scores reflect factors for which they were not responsible, as the scores were based on the performance of students Wagner and Braeuner did not teach. Stakes were fairly high, reliability was low, and the teachers bore little responsibility for the outcomes. So, per our account, they were wronged.

Houston Fed of Teachers v. HISD. In 2012, the Houston Independent School District ("Houston schools”) began using a similar proprietary IVA (EVAAS) to evaluate teachers. Houston Schools had the "aggressive goal of 'exiting' 85% of teachers with 'ineffective' EVAAS ratings." (251 F. Supp. 3d 1168, 1174-75). And in the first three years using EVAAS, Houston Schools "exited" between 20% and 25% of the teachers rated ineffective. Moreover, the district court determined that the EVAAS scores were the sole basis for those actions (251 F. Supp. 3d 1168, 1175).

As in *Wagner*, the *Houston* court determined that the teachers did not have their substantive due process rights violated because use of EVAAS cleared the low rational basis standard. However, the court determined that the teachers' *procedural* due process rights were infringed. Because the system is proprietary, there was no meaningful way for teachers to ensure that their individual scores were calculated correctly. The court noted that there were apparently no mechanisms to correct basic clerical and coding errors. And where such mistakes did occur in a teacher's score, Houston Schools refused to correct them because the correction process disrupts the analysis. In response to a "frequently asked question" HISD states:

> Once completed, any re-analysis can only occur at the system level. What this means is that if we change information for one teacher, we would have to run the analysis for the entire district, which has two effects: one this would be very costly for the district, as the analysis itself would have to be paid for

> again; and two, this re-analysis has the potential to change all other teachers' reports (emphasis in original).[12]

That last point is worth stressing. Each teacher's individual score is dependent on all other teachers' scores. So, a mistake for one teacher's score affects all others' scores. As the court states, "this interconnectivity means that the accuracy of one score hinges upon the accuracy of all" 251 F. Supp. 3d 1168, 1178.

The moral foundations of the teachers' complaints will by now be obvious. The stakes here—i.e. losing one's job and having one's professional image tarnished—are high. EVAAS is unreliable, having what the court called a "house-of-cards fragility" (1178). And that unreliability is due to factors for which teachers are not responsible, "ranging from data-entry mistakes to glitches in the code itself" (1177). Hence, teachers could not reasonably endorse being evaluated under such a system.

Loomis v. WI. Our argument that algorithmic systems conflict with persons' autonomy extends to the *Loomis* case. To begin, COMPAS is only moderately reliable. Researchers associated with Northpointe assessed COMPAS as being accurate in about 68% of cases (Brennan, Dietrerich, and Ehret, 2009)**.**[13]

More important is that COMPAS incorporates numerous factors for which defendants are not responsible. COMPAS takes a number of data points about a defendant's criminal behavior, history, beliefs, and job skills, and generates a series of risk scales. These include pretrial release risk (likelihood that defendant will fail to appear in court or have a new felony

[12] http://static.battelleforkids.org/documents/HISD/EVAAS-Value-Added-FAQs-Final-2015-02-02.pdf.

[13] A study by ProPublica found that prediction failure was different for white and black defendants, such that white defendants labeled lower risk were more likely to re-offend than black defendants with a similar label, and black defendants labeled higher risk were less likely to re-offend than white defendants labeled higher risk (Angwin et al, 2016).

arrest if released prior to trial); risk of general recidivism (whether defendant will have subsequent, new offenses); and risk of violent recidivism (Northpointe 2015, 27-28). Among the factors that COMPAS uses to assess these risks are current and pending charges, prior arrests, residential stability, employment status, community ties, substance abuse, criminal associates, history of violence, problems in job or educational settings, and age at first arrest. (Northpointe 2015, 24). Incorporating these factors into a proprietary algorithm, COMPAS generates bar charts corresponding to degree of risk. According to Northpointe, "[b]ig bars, bad—little bars, good," at least as a first gloss (Northpointe 2015, 4). Loomis's COMPAS report indicated that he presented a high risk of pretrial recidivism, general recidivism, and violent recidivism. (¶16)

Regardless of just how well COMPAS's big and little bars reliably reflect recidivism risk, defendants are not responsible for some of the factors that affect those bars. So, while Loomis did commit the underlying conduct and was convicted of prior crimes, COMPAS incorporates factors for which defendants are not responsible.[14] For example, the questionnaire asks about the age at which one's parents separated (if they did), whether one was raised by biological, adoptive, or foster parents, whether a parent or sibling was ever arrested, jailed, or imprisoned, whether a parent or parent-figure ever had a drug or alcohol problem, and whether one's neighborhood friends or family have been crime victims.[15]

[14] Drawing on factors for which one isn't responsible is compatible with a range of theories of punishment. Such factors may help determine a sentence—whether one is even arrested, moral luck, how well punishment deters crime, and so forth. But our view is not that only factors for which one is responsible may contribute to sentencing decisions. Rather, our view is that, as such factors increase, it becomes more difficult for an agent to endorse such a system.

[15] Other questions pertain to matters for which defendants' responsibility is less clear: how often one has had barely enough money to get by, whether one's friends use drugs, how often one has moved in the last year, and whether one has ever been suspended from school.

Finally, the use of COMPAS in *Loomis* is high-stakes. Incarceration is the harshest form of punishment that the state of Wisconsin can impose. This is made vivid by comparing the use of COMPAS in *Loomis* with its specified purposes. COMPAS is built to be applied to decisions about the type of institution in which an offender will serve a sentence (e.g., lower or higher security), the degree of supervision (e.g., from probation officers or social workers), and what systems and resources are appropriate (e.g., drug and alcohol treatment, housing, and so forth). Indeed, Northpointe warns against using COMPAS for sentencing, and Loomis's presentence investigation report specifically stated the COMPAS report should be used "to identify offenders who could *benefit from interventions and to target risk factors that should be addressed during supervision*" (¶16, emphasis added).

When the system is used for its intended purposes—identifying ways to mitigate risk of recidivism of persons under state supervisions, the stakes are much lower.[16] Hence, it is more plausible that Loomis (or any offender) could reasonably endorse such a system.

4. Knowing Where One Stands

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 (Pasquale 2016). But it is worth asking why transparency is important. Transparency may be important for instrumental purposes, and in the case of public use of algorithms,

[16] Northpointe describes COMPAS's scope as follows: "Criminal justice agencies across the nation use COMPAS to inform decisions regarding the placement, supervision and case management of offenders." (Northpointe 2015, p. 1).

transparency may be important for accountability (Powles 2017).[17] Our view is that transparency is also integral for respecting agency.[18] To see why, consider the following.

As we argued in section 2, agents are autonomous only if they are able to incorporate 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. So, it is a failure of respect for autonomy to prevent agents from exercising their autonomy without good reason.

Incorporating one's values into important facets of one's life requires that one have access to relevant information. This is on account of two distinct aspects of agency (for a similar division of aspects of our agency and discussion, see Smith 2013). One aspect of agency is the ability to take actions that realize one's values. Call this 'practical agency'. 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 organization functions, 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.

[17] There's a further question about what should be transparent. That is, should underlying code be transparent, should it merely be open to be audited, should the uses of the algorithm be made clear, and so forth. Which of these aspects of transparency is important will turn on the underlying moral justification for transparency.

[18] One recent account of transparency focuses the appropriate function of transparency. Sloan and Warner (2018) argue that algorithms are transparent with respect to consumers where consumers are readily able to ascertain risks and benefits of the system. This bridges both autonomy views (because it addresses the degree to which consumers may exercise agency over decisions) and consequentialist views (because the criteria considers only consequences of algorithm use, and not other moral factors).

The importance of transparency does not solely depend on agents' abilities to use information to act. A second aspect of agency is the ability to *understand* important facets of one's life. Call this 'cognitive agency'. As Thomas Hill, Jr. has persuasively 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 (Hill, 1984). Even a benevolent lie that spares another's feelings can be an affront because it thwarts that person's ability to understand her situation. 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 (Rubel 2007).

Since autonomy requires having information relevant to one's life, respecting autonomy requires not denying agents that information. Algorithmic decision systems are often not built to be transparent (Pasquale 2016). As we show next, this denies agents information to which they have a right.

4.1 Legal Disputes

Houston Fed of Teachers v. HISD. A central complaint in this case was that EVAAS was "too vague to provide notice to teachers of how to achieve higher ratings and avoid adverse employment consequences" (251 F. Supp. 3d at 1173). Our analysis helps make moral sense of this complaint.

Knowing how EVAAS works enables teachers to make better decisions in important facets of their lives, and hence to exercise practical agency. It may, for example, help them bring their instruction in line with Houston Schools’ goals, thereby making their employment more secure. Or it could give teachers grounds for impeaching EVAAS through legal action, appeal to the public, or collective bargaining.

Yet, before the case was litigated, the Houston teachers were unable to gain this understanding. As the court points out, "SAS [EVAAS's developer] treats these algorithms and software as trade secrets, refusing to divulge them to either [Houston Schools] or the teachers themselves" (251 F. Supp. 3d at 1177). Even if EVAAS is reliable, respecting teachers requires enabling them to understand its decisions, and if it is unreliable, it warrants an appeal for obvious reasons. Teachers may have no such legal right.[19] But this is a moral argument, not a legal one.

Teachers also have a claim to understand EVAAS on grounds of *cognitive* agency. Losing one's job on account of being "ineffective" is highly significant to the teachers who were exited on the basis of their EVAAS scores. Simply knowing that EVAAS was flawed may help teachers maintain a sense of self-worth in the face of such firings.[20] And, hence, they have a claim on the basis of cognitive agency to understand that event.

Loomis v. WI. 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.

In *Gardner v. Florida* (430 U.S. 349), 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

[19] In fact, a court has determined that in low-stakes cases they do not See *Trout v. Knox County Board of Education* (163 F. Supp. 3d 492).

[20] This point also applies to *Wagner v. Haslam.* Wagner and Braeuner's scores—which, recall, were largely based on a school-wide composite score—simply can't be understood as a plausible measurement of their teaching ability. This fact may frustrate them and give them cause to take action against further use of TVAAS, but it is also likely to mitigate any feelings of ineptitude that may have been brought on by a low assessment.

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 (see ¶51), 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 the same system), Loomis argued that he was denied the opportunity to refute or explain his results.

The *Loomis* 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 COMPAS being proprietary. ¶¶54-56. 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 reportss using COMPAS should include a number of warnings about its limitations

We can offer two distinct ways to morally underwrite Loomis’ complaint: one based in practical agency and another based in cognitive agency. Loomis faced a number of decisions about what to do in response to 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 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.

Regardless of the concerns based on practical agency, Loomis has a claim to better understand the process that generated his risk score and facts about whether his risk score fairly represents him. 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. That plausibly includes access to both proprietary and audit information. And his claim based on cognitive agency does not turn on whether access to such information would aid his case.

5. Herding

There is yet another way in which algorithmic systems conflict with agency. They can aggregate individuals' interests rather than regarding each group-member's interests as separate. Call this aggregation of interests "herding."

In Rawls's terms, our criticism is that algorithmic systems fail to "take seriously the distinction between persons" (1971/1999, 24). Rawls's target was classical utilitarianism, which aggregates each person's interests into the interests of a single representative agent, and which uses the principle of individual rational choice maximization as a principle of social welfare. The methodological purpose of the representative agent is to carry out "the required organization of the desires of all persons into one coherent system of desire" (24).

As Rawls points out, we do not in fact all share a single, unified system of desire. He argues that the utilitarian decision procedure makes an ontological mistake about the fundamental individuality (or "separateness") of persons. Here, we offer the same sort of argument about algorithmic decision-making systems in general. Since algorithmic systems are ubiquitous across contemporary life, they have a capacity to manipulate those they manage, and they will not *automatically* attend to the separateness of persons, we should expect the herding problem to be pervasive.[21]

And it is. Uber, for instance, adjudicates customer complaints in part by examining whether the driver deviates from the route suggested by Uber's mapping software. However, the software routes drivers so as to efficiently balance Uber's incentives, which are quite different than the personal incentives of their customers and their drivers (Calo and Rosenblat 2017, 1669).

With respect to any geographical space it wishes to serve, Uber has two distinct and competing incentives. One is to *explore* that space as fully as possible in order to gain an exhaustive understanding of it. The other is to *exploit* that understanding for the purposes of efficient routing. To balance these two incentives, Uber employs a "multi-arm bandit algorithm" (*ibid.*), which is a method of information processing also known as "ant colony optimization."

When ants search a space for food or other resources, they leave behind trails of pheromones. These trails embed knowledge in the space for other ants in the colony to exploit.

[21] For instance, the scarcity of healthcare resources has led to the use of cost-effectiveness analysis, which aims to ration those resources so as to optimize well-being across the relevant population. This form of analysis is attractive at the level of the population, but it raises a variety of concerns at the level of the individual, because it achieves its goal of optimization precisely through herding.

In the case of both Uber drivers and ants, the well-being of the "explorers" is secondary to the well-being of the herd that exploits their exploratory work. This is not as troubling with respect to ants as it is with respect to human workers, because ants are not governed by substantive individual aims.[22] Moreover, they directly share in the spoils of the colony's collective exploration. In contrast, Uber drivers carry out the exploratory task within Uber's scheme, but at best share in the spoils indirectly. As Calo and Rosenblat write, “assuming Uber is training its own systems on the limitless driver data to which it has access, Uber participants may be unwittingly training their replacements” (*ibid.*).

Uber fails to treat the interests of its drivers as fundamentally separate from its corporate interests. It manipulates its drivers to follow the collective scheme using insights from behavioral psychology. The *New York Times* reported that some local Uber managers "went so far as to adopt a female persona for texting drivers," because "uptake was higher when they did." Uber also gamifies its drivers' user interfaces, to manage driver perceptions of both how well they have done, and how well they *could* do if they only kept going a little longer. As long as the company's most important assessment metrics are growth and volume, it will have "an incentive to make wringing more hours out of drivers a higher priority than the drivers’ bottom line," and "an incentive to obtain these hours as cheaply as possible" (Schieber 2017).

Uber's failures do not necessarily demonstrate that it has acted wrongly. Algorithmic systems operate by generalizing about individuals, and Uber may not have a moral obligation to treat everyone as separate persons. Uber is not making an ontological mistake in herding their

[22] When they are removed from colony structures, they simply follow a random walk.

drivers, because it is not grounding individual responsibility in facts about collective performance.

5.1 Legal Disputes

In our target cases collective performance *does* appear to be playing a grounding role. In *Wagner*, the plaintiffs were teachers in non-tested subjects, and had been evaluated as individuals based on the aggregated scores of students in tested subjects. Before 2013-14, the two teachers had received high individual scores. However, when their schools' aggregated scores fell, the teachers were denied performance bonuses and consideration for tenure. The mistakes were ontological: the teachers' fundamentally separate actions (that is, their individual performances) had been conflated with the aggregated performance of their schools.

Loomis raises the same concern. One of Loomis's central arguments is that his right to an individualized sentence was violated. COMPAS works by analyzing data from a large number of previous cases. It identifies factors that have been shown to correlate with new offenses, with new offenses pre-trial, and with violence. By comparing a new offender to the information in the COMPAS databases, the tool creates a risk score. Loomis argued that this is tantamount to non-individuated sentencing. He argued, in other words, that he was sentenced as a member of the group classified as 'high risk' ("big bars bad") rather than as an individual.

It is worth noting that the Wisconsin Supreme Court agreed that basing a sentence on a COMPAS score alone would indeed violate a right to an individual sentence. It affirmed Loomis's sentence on the grounds that the circuit court did not rely on the score alone. Rather, the judge explicitly considered individual factors about Loomis, including the read-in charges[23] and Loomis's history while under supervision.[24] In other words, while COMPAS may well form a sound basis for statistical generalizations about the behavior of populations, individuals have a right to an individualized sentence. The Wisconsin court recognizes this as a legal right. That legal right is underwritten by a moral obligation to recognize a distinction between persons, which is in turn underwritten by respect for persons.

Wagner, *Trout v. Knox County Board of Education*, and *Loomis* help illustrate how problems raised by algorithmic herding resist legal remedy. In *Wagner*, the court acknowledged "significant problems with the pace and nature" of TVAAS implementation but determined that it cleared the low hurdle of rational basis review (112 F.Supp. 3d at 698). In *Trout*, the Eastern District of Tennessee determined that teachers challenging IVA assessments lacked a property interest sufficient to trigger due process protections in the first place. Although the state supreme court in *Loomis* set some limits on the permissible use of automated risk assessment tools, it determined that use of those tools is within the circuit court's discretion.

So, courts do seem to recognize some of the hazards that are associated with algorithmic herding. However, they are reluctant to overturn conclusions of algorithmic systems. This limits the ability of people subject to those systems to be treated as individuals.

[23] When charges are "read in" the judge assumes that the basic facts of the case are true and that the defendant was involved in the underlying conduct. ¶20. The trial court viewed those read-in charges as a "serious, aggravating factor," ¶20.

[24] "[T]he record reflects that although the circuit court referenced the risk assessment at sentencing, the court essentially gave it little or no weight." ¶105.

6. Agency Laundering

In each of the previous sections we have explained how a person or entity may use algorithmic decision-making systems in a way that conflicts with others' agency or fails to respect them as persons. A different issue concerns not the agency of those who are subjects of algorithms but the agency of those who deploy them. Using an algorithm to make decisions can allow an agent to distance herself from morally suspect actions by attributing morally relevant characteristics to the algorithm.

Call it "agency laundering."[25]

To understand the idea of agency laundering, it's useful to start by comparing it money laundering.[26] People who have a great deal of cash that they wish to hide (for example to hide an illegal enterprise or to avoid tax liability) may mix that cash with other money (perhaps from legitimate sources) to avoid suspicion. Of course decisions are not the same as cash, and our argument does not turn on the analogy. Our point is instead that, like money laundering, agency laundering involves obscuring the source of something dubious by mixing it with something similar, but seemingly above-board.

[25] The concept of agency laundering is developed more fully in Rubel, Castro, and Pham (2019a) and (2019b). The version articulated here follows more closely the condensed treatment in (2019b). The point in this paper is that considerations of agency and autonomy are key in understanding moral questions underlying algorithmic systems rather than a full explication of agency laundering.
Agency laundering is distinct from some superficially similar concepts. For example, Barocas and Selbst (2016) describe "masking," or the intentional use of algorithmic systems to obfuscate discrimination. Laundering could accompany (or follow) masking, but a key component of laundering is that the person with authority ascribes morally relevant qualities to the algorithm. Masking does not require this. Mattias (2004) discusses the idea of a responsibility gap, which refers to the inadequacy of traditional ascription of responsibility in the context of automated computational artifacts. The belief in a responsibility gap may enable agency laundering, but agency laundering entails that some agent actually has responsibility.

[26] 18 U.S. Code § 1956 - Laundering of monetary instruments.

Consider “consultant.” Suppose that a business proprietor believes she can increase her profits by laying off older, better-compensated employees and replacing them with contractors. The proprietor hires a consultant to evaluate her business’s practices knowing full well that the consultant will recommend laying off the older employees. She then does as the consultant recommends.

The proprietor could have simply fired her employees and hired contractors. Instead, she hired a consultant to recommend what she wanted to do in the first place. When employees hear about the layoffs, the proprietor can point to the consultant report as the rationale for the layoffs. Her actions ensured a particular result, but by hiring the consultant she launders her agency by obscuring her role in the decision. Thus, the consultant’s research appears to be the relevant decision-maker, even though it was the proprietor all along.

The morally important facet of consultant, though, is *not* whether it is morally justifiable for the proprietor to fire her employees (perhaps so, perhaps not). Rather, it is that the proprietor had de jure and de facto authority to do so and appeared to hand the reasoning behind that decision to a separate entity. By hiring the consultant, she implied that the consultant was disinterested, competent to evaluate business practices, would weigh all relevant evidence judiciously, and had the power to return a report that did not merely reflect the proprietor’s wishes. But those implications were not all true—*ex hypothesi*, the proprietor justifiably believed that the consultant would return the results that the proprietor wanted. The proprietor is able to obscure her role in determining how she would increase profits and launder her agency.

So, an agent, *a*, *launders her agency* where:

1. *a* has de facto and de jure authority with respect to φ-ing (where to φ is an action);
2. *a* gives *b* (some process, person, or entity) de facto practical authority with respect to φ-ing;
3. *a* ascribes (implicitly or explicitly) morally relevant qualities to *b*'s conclusions (e.g., relevance, neutrality, reliability);
4. *a* thereby obscures her de facto and de jure authority for φ-ing.

Understanding agency laundering helps to make sense of underlying moral issues in some of the legal disputes that we have discussed.

6.1. Legal disputes

Houston Fed of Teachers v. HISD. HISD launders its responsibility for firing teachers. It has de facto and de jure authority to hire, fire, and promote teachers, and it is morally responsible for that process. It defers to EVAAS in making those determinations, hence giving EVAAS de facto practical authority. HISD also ascribes morally relevant qualities to EVAAS, repeatedly referring to its rigor, complexity, and reliability.[27] The fact, if it is, that EVAAS accurately measures student progress over time does not entail that EVAAS measures the extent to which teachers are responsible for that progress (or lack thereof).[28] By referring to one good thing that EVAAS does, HISD obscures the fact that they are firing teachers based on measures for which the teachers are not responsible.

HISD's claims that EVAAS is rigorous and complex further obscure their responsibility. In a white paper on EVAAS, SAS notes "[r]ather than focus on simplicity of calculation, SAS EVAAS models prioritize reliability of analysis and then focus on ease of interpretation and ease of

[27] See Defendant response to complaint, ¶¶ 27-39, *Houston Federation of Teachers v. HISD*, No. 4:14-cv-01189.

[28] EVAAS presupposes that "socioeconomic and demographic influences persist over time," and that "those influences are represented and accounted for in the student‘s data." This, they claim, enables them to use "each student as her or her own control" (Defendant response to complaint, ¶¶ 33, *Houston Federation of Teachers v. HISD*, No. 4:14-cv-01189). This raises a concern about EVAAS's reliability. Socioeconomic and demographic influences *do* change over time.

usage" (Sanders et al., 2009). SAS's choice to prioritize reliability over understandability inoculates EVAAS from scrutiny. This choice is a form of obfuscation.

Wisconsin v. Loomis. *Loomis* is where our conception of algorithmic agency laundering has some bite. The Wisconsin Supreme Court's reasoning in the case diminishes a circuit court's ability to launder its agency by using tools like COMPAS.

A trial court launders its agency where:

1. It has de facto and de jure authority with respect to sentencing;
2. It gives COMPAS de facto practical authority with respect to sentencing;
3. It ascribes morally relevant qualities to COMPAS's conclusions;
4. It thereby obscures its de facto and de jure authority for sentencing.

The trial court certainly has de facto and de jure authority for sentencing. The trial court gave COMPAS some degree of de facto authority in sentencing. The judge referenced the COMPAS risk scores, but he also explained the relevance of the read-in charges and Loomis's history under supervision. Regarding the third condition, it does appear that the trial court implicitly ascribes morally relevant qualities to COMPAS (e.g., that it is accurate and neutral).

The crucial factor in considering whether *Loomis* involves agency laundering is criterion four. In its decision, the Wisconsin Supreme Court's requires that use of COMPAS be supported by other factors that are independent of the algorithm. Specifically, the court is explicit that a COMPAS finding may not be the sole determining factor in sentencing and supervision:

> We determine that because the circuit court explained that its consideration of the COMPAS risk scores was supported by other independent factors, its use was not determinative in deciding whether Loomis could be supervised safely and effectively in the community. Therefore, the circuit court did not erroneously exercise its discretion. ¶9

The court goes on to impose several conditions for any use of COMPAS for sentencing purposes. It requires that courts weigh all relevant factors in sentencing. (¶74) It also prohibits the use of scores in determining whether to incarcerate a person and in determining the length or severity of their sentence. ¶¶88-98. Finally, the court requires that any presentence investigation report using COMPAS contain a warning about its limitations.

The supreme court's interpretation is that the trial court made only limited use of COMPAS and incorporated other factors in doing so, and it imposed limits on the use of COMPAS and similar algorithmic systems. This addresses both condition three and condition four. With respect to condition three, the court's opinion seeks to ensure that trial courts do *not* ascribe certain morally salient characteristics to COMPAS, viz., that it is a wholly reliable tool for assessing risk. Further, under the supreme court's understanding, sentencing decisions are wholly the trial court's responsibility, and relying on the COMPAS algorithm cannot distance the trial court from that responsibility.

Hence, relying on an algorithm like COMPAS could allow courts and others in the criminal justice system to launder their agency. However, the *Loomis* decision appears tailored to avoid that, and the trial court's decision does not appear to be an instance of laundering.

7. Conclusion and Caveats

The literature on algorithmic decision systems points to ways that such systems can be unfair, can cause harm, discriminate, thwart accountability, or undermine autonomy by being manipulative. Our arguments here do not discount those moral issues. We have made the case that the moral and legal salience of algorithmic systems requires attention to broader issues of agency, autonomy, and respect for persons. Algorithmic systems may govern behavior in ways that an agent cannot reasonably endorse. They may deny an agent information to which she is entitled. They may fail to respect boundaries between persons. And they may be deployed to launder agency.

One issue worth addressing here is whether our arguments identify problems that are unique to algorithmic decision systems. The moral concerns we describe can exist in any kind of decision system. Decision processes that rely on (inter alia) committee or bureaucracy can be processes that individuals cannot reasonably endorse, can be opaque, can fail to treat people as separate individuals, and can launder agency. Our task here, though, has been to examine several types of moral concern in decision-making and how those relate to algorithmic decision processes. That those same concerns apply beyond algorithms shows that the root moral concerns are deeper.

Moreover, there are features of algorithmic decision-making that will make some of the concerns we describe particularly acute. First, because of a kind of aura that surrounds mathematical models, people end up trusting them disproportionately (O'Neil 2016; Zarsky 2016; Citron 2008). Second, because such systems are difficult to understand and many believe them to be difficult to understand, people may be reluctant to criticize them. Of course, humans, committees, and bureaucracies are difficult to understand as well, but we may have an intuitive grasp of the kinds of faults in reasoning that they exhibit (motivated reasoning, groupthink, various cognitive heuristics). Lastly, in addition to being complex, algorithmic systems are often proprietary and protected by intellectual property rights, and hence enjoy legal protections that other systems do not (Pasquale 2016).

References

Angwin, Julia, and Jeff Larson. “The Tiger Mom Tax: Asians Are Nearly Twice as Likely to Get a Higher Price from Princeton Review,” September 1, 2015. https://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review.

Angwin, Julia, Jeff Larson. “Machine Bias.” ProPublica, May 23, 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Angwin, Julia, Madeleine Varner. “Facebook Enabled Advertisers to Reach ‘Jew Haters.’” Text/html, September 14, 2017. https://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters.

Barocas, Solon and Andrew Selbst. “Big Data’s Disparate Impact.” *California Law Review* 104, no. 3 (2016): 671-732.

Brennan, Tim, William Dieterich, and Beate Ehret. “Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System.” *Criminal Justice and Behavior* 36, no. 1 (2009): 21–40. https://doi.org/10.1177/0093854808326545.

Calo, Ryan, and Alex Rosenblat. “The Taking Economy: Uber, Information, and Power.” *Columbia Law Review* 117, no. 6 (2017): 1623-1690.

Citron, Danielle. “Data Mining for Juvenile Offenders.” Accessed April 21, 2018. https://concurringopinions.com/archives/2010/04/data-mining-for-juvenile-offenders.html.

Citron, Danielle Keats. “Technological Due Process.” *Washington University Law Review* 85, no. 6 (2008): 1249–1314.

Citron, Danielle, and Frank Pasquale. “The Scored Society: Due Process for Automated Predictions.” *Faculty Scholarship*, January 1, 2014. http://digitalcommons.law.umaryland.edu/fac_pubs/1431.

Danaher, John. “The Threat of Algocracy: Reality, Resistance and Accommodation.” *Philosophy & Technology* 29, no. 3 (2016): 245–68.

Dixon, Pam, and Bob Gellman. “The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future.” World Privacy Forum, 2014.

Engel, Pascal. “Is Epistemic Agency Possible?” *Philosophical Issues* 23, no. 1 (2013): 158–78. https://doi.org/10.1111/phis.12008.

Equivant. “Northpointe Suite Automated Decision Support,” 2018. http://www.equivant.com/assets/img/content/Northpointe_Suite_2.pdf.

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin’s Press, 2018.

Hill, Jr., Thomas. “The Kantian Conception of Autonomy.” In *The Inner Citidel: Essays on Individual Autonomy*, 91–105, 1989.

Hill, Thomas E. “Autonomy and Benevolent Lies.” *The Journal of Value Inquiry* 18, no. 4 (December 1, 1984): 251–67. https://doi.org/10.1007/BF00144766.

Isenberg, Eric, and Heinrich Hock. "Measuring School and Teacher Value Added in DC, 2011-2012 School Year." Mathematica Policy Research, Inc, August 31, 2012. https://eric.ed.gov/?id=ED565712.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor, in *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Cambridge, UK: Cambridge University Press, 1996.

Kant, Immanuel. *The Metaphysics of Morals*. Translated by Mary Gregor, in *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy*. Cambridge, UK: Cambridge University Press, 1996.

Lanzing, Marjolein. "'Strongly Recommended' Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies." *Philosophy & Technology* 32, no. 3 (2019): 549-568.

Mackenzie, Catriona, and Natalie Stoljar, eds. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. 1 edition. New York: Oxford University Press, 2000.

Meyers, Diana T. "Personal Autonomy and the Paradox of Feminine Socialization." *The Journal of Philosophy* 84, no. 11 (1987): 619–28.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society* 3, no. 2 (2016): 1-21. https://doi.org/10.1177/2053951716679679.

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

Northpointe, Inc. "Practitioner's Guide to COMPAS Core," March 19, 2015. http://www.northpointeinc.com/files/technical_documents/Practitioners-Guide-COMPAS-Core-_031915.pdf.

Nussbaum, Martha C. *Sex and Social Justice*. Oxford University Press, 1999.

Olson, Dustin. "A Case for Epistemic Agency." *Logos & Episteme* 6, no. 4 (November 1, 2015): 449–74. https://doi.org/10.5840/logos-episteme20156435.

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

Oremus, Will, and Bill Carey. "Facebook's Offensive Ad Targeting Options Go Far beyond 'Jew Haters.'" *Slate*, September 14, 2017. http://www.slate.com/blogs/future_tense/2017/09/14/facebook_let_advertisers_target_jew_haters_it_doesn_t_end_there.html.

Oshana, Marina A. L. "Personal Autonomy and Society." *Journal of Social Philosophy* 29, no. 1 (2018): 81–102. https://doi.org/10.1111/j.1467-9833.1998.tb00098.x.

Parfit, Derek. *On What Matters: Two-Volume Set*. Oxford University Press, 2011.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press, 2016.

Peck, Don. “They’re Watching You at Work.” *The Atlantic*, December 2013. https://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/.

Pham, Adam and Clinton Castro. 2019. “The Moral Limits of the Market: The Case of Consumer Scoring Data.” *Ethics and Information Technology* 21, no. 2: 117-126. DOI 10.1007/s10676-019-09500-7.

Powles, Julia. “New York City’s Bold, Flawed Attempt to Make Algorithms Accountable.” *The New Yorker*, December 21, 2017. https://www.newyorker.com/tech/elements/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable.

Quick, Kimberly. “The Unfair Effects of IMPACT on Teachers with the Toughest Jobs.” *The Century Foundation* (blog), October 16, 2015. https://tcf.org/content/commentary/the-unfair-effects-of-impact-on-teachers-with-the-toughest-jobs/.

Rawls, John. *A Theory of Justice*. Cambridge, Mass.: Belknap Press of Harvard University Press, 1999.

Raz, Joseph. *The Morality of Freedom*. Oxford University Press, 1988.

Rubel, Alan. 2007. “Privacy and the USA Patriot Act: Rights, the Value of Rights, and Autonomy.” *Law & Philosophy* 26, no. 2: 119–59.

Rubel, Alan, Clinton Castro, and Adam Pham. 2019a. “Agency Laundering and Information Technologies.” *Ethical Theory & Moral Practice* 22, no. 4: 1017–1041 (2019). DOI 10.1007/s10677-019-10030-w

Rubel, Alan, Clinton Castro, and Adam Pham. 2019b. “Agency Laundering and Algorithmic Decision Systems.” In: Taylor N, Christian-Lamb C, Martin M, Nardi B (eds.) Proceedings of the 2019 iConference, *Information in Contemporary Society (Lecture Notes in Computer Science)*. Springer Nature, pp 590-598.

Sanders, William, S. Paul Wright, June C. Rivers, and Jill G. Leandro. “Addressing Common Concerns about Value-Added Modeling,” November 2009. http://www.education.pa.gov/Documents/K-12/Assessment%20and%20Accountability/PVAAS/Methodology%20and%20Research%20Materials/Addressing%20Common%20Concerns%20about%20Value-Added%20Modeling.pdf.

Scanlon, T. M. *What We Owe to Each Other*. Revised edition. Cambridge, Mass.: Belknap Press, 2000.

Shiffrin, Seana Valentine. “Paternalism, Unconscionability Doctrine, and Accommodation.” *Philosophy & Public Affairs* 29, no. 3 (2000): 205–50.

Sloan, Robert and Richard Warner. “When Is an Algorithm Transparent?: Predictive Analytics, Privacy, and Public Policy,” *IEEE Security & Privacy* 16, no. 3 (2018): 18-25. https://doi.org/10.1109/MSP.2018.2701166

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

Sosa, Ernest. *Judgment and Agency*. Oxford, New York: Oxford University Press, 2015.

Strauss, Valerie. “D.C. Teacher Tells Chancellor Why IMPACT Evaluation Is Unfair.” *Washington Post*, August 16, 2011, sec. Local. https://www.washingtonpost.com/blogs/answer-sheet/post/dc-teacher-tells-chancellor-why-impact-evaluation-is-unfair/2011/08/15/gIQA0yhBIJ_blog.html.

Superson, Anita. “Deformed Desires and Informed Desire Tests.” *Hypatia* 20, no. 4 (July 20, 2018): 109–26. https://doi.org/10.1111/j.1527-2001.2005.tb00539.x.

Susser, Daniel, Beate Roessler, and Helen Nissenbaum, “Online Manipulation: Hidden Influences in a Digital World.” *Georgetown Law Technology Review* 4, no. 1 (2019): 1-45.

Thomson, Judith. “Liability and Individualized Evidence.” *Law and Contemporary Problems* 49, no. 3 (1986): 199–219.

Trout v. Knox County Board of Education (163 F. Supp. 3d 492, E.D. Tenn. 2016)

Turow, Joseph. *The Aisles Have Eyes: How Retailers Trackt Your Shopping, Strip Your Privacy, and Define Your Power*. New Haven: Yale University Press, 2017.

Turque, Bill. “More than 200 D.C. Teachers Fired.” *Washington Post*, July 15, 2011, sec. Local. https://www.washingtonpost.com/blogs/dc-schools-insider/post/more-than-200-dc-teachers-fired/2011/07/15/gIQADnTLGI_blog.html.

Walsh, Elias, and Dallas Dotter. “Longitudinal Analysis of the Effectiveness of DCPS Teachers.” Mathematica Policy Research Reports. Mathematica Policy Research. Accessed April 21, 2018. https://ideas.repec.org/p/mpr/mprres/65770df94dde4573b331ce1cb33a9e07.html.

Wisconsin v. Loomis (Wisconsin Supreme Court 2016).

Yeung, Karen. “’Hypernudge’: Big Data as a Mode of Regulation by Design.” *Information, Communication & Society* 21, no. 1 (2017): 118-136.

Zarsky, Tal. “The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness in Automated and Opaque Decision Making.” *Science, Technology, & Human Values* 41, no. 1 (2016): 118–32. https://doi.org/10.1177/0162243915605575.