Journal of Business Ethics (2019) 160:835–850
https://doi.org/10.1007/s10551-018-3921-3

ORIGINAL PAPER

Ethical Implications and Accountability of Algorithms

Kirsten Martin[1]

Received: 3 July 2017 / Accepted: 5 April 2018 / Published online: 7 June 2018

Abstract
Algorithms silently structure our lives. Algorithms can determine whether someone is hired, promoted, offered a loan, or provided housing as well as determine which political ads and news articles consumers see. Yet, the responsibility for algorithms in these important decisions is not clear. This article identifies whether developers have a responsibility for their algorithms later in use, what those firms are responsible for, and the normative grounding for that responsibility. I conceptualize algorithms as value-laden, rather than neutral, in that algorithms create moral consequences, reinforce or undercut ethical principles, and enable or diminish stakeholder rights and dignity. In addition, algorithms are an important actor in ethical decisions and influence the delegation of roles and responsibilities within these decisions. As such, firms should be responsible not only for the value-laden-ness of an algorithm but also for designing who-does-what within the algorithmic decision. As such, firms developing algorithms are accountable for designing how large a role individual will be permitted to take in the subsequent algorithmic decision. Counter to current arguments, I find that if an algorithm is designed to preclude individuals from taking responsibility within a decision, then the designer of the algorithm should be held accountable for the ethical implications of the algorithm in use.

Keywords Algorithms · AI · Ethics · Fairness · Artificial intelligence · Big Data · Accountability · STS · Technology

The New York Times

Q: Whose responsibility is it to ensure that algorithms or software are not discriminatory?

A: This is better answered by an ethicist.

Cynthia Dwork, computer scientist at Microsoft Research, Gordon McKay Professor of Computer Science at Harvard University.

> Rodríguez was just sixteen at the time of his arrest, and was convicted of second-degree murder for his role in an armed robbery of a car dealership that left an employee dead. Now, twenty-six years later, he was a model of rehabilitation. He had requested a transfer to Eastern, a maximum-security prison, in order to take college classes. He had spent four and a half years training service dogs for wounded veterans and eleven volunteering for a youth program. A job and a place to stay were waiting for him outside. And he had not had a single disciplinary infraction for the past decade… Yet, last July, the parole board hit him with a denial. It might have turned out differently but, the board explained, a computer system called COMPAS had ranked him "high risk." Neither he nor the board had any idea how this risk score was calculated; Northpointe, the for-profit company that sells COM-

Editors at the Journal of Business Ethics are recused from all decisions relating to submissions with which there is any identified potential conflict of interest. Submissions to the Journal of Business Ethics from editors of the journal are handled by a senior independent editor at the journal and subject to full double-blind peer-review processes.

✉ Kirsten Martin
martink@gwu.edu

[1] George Washington University, Washington, DC, USA

PAS, considers that information to be a trade secret. (Wexler 2017).[1]

Algorithms silently structure our lives. Not only in determining your search results and the ads you see online, algorithms can also predict your ethnicity (Garfinkel 2016), who is a terrorist (Brown 2016), what you will pay (Angwin et al. 2016b), what you read (Dewey 2016), if you get a loan (Kharif 2016), if you have been defrauded (Nash 2016), if and how you are targeted in a presidential election (O'Neil 2016), if you are fired (O'Neil 2016), and most recently, if you are paroled and how you are sentenced (Angwin et al. 2016a; Wexler 2017). The insights from Big Data do not come from an individual looking at a larger spreadsheet. Algorithms sift through data sets to identify trends and make predictions. While the size of data sets receives much of the attention within the Big Data movement,[2] less understood yet equally important is the reliance on better, faster, and more ubiquitous algorithms to make sense of these ambiguous data sets. Large data sets without algorithms just take up space, are expensive to maintain, and provide a temptation for hackers. Algorithms make data sets valuable.

The benefits of algorithms parallel the many benefits of Big Data initiatives: we have more tailored news, better traffic predictions, more accurate weather forecast, car rides when and where we want them. And yet, we continue to see headlines about algorithms as unfairly biased and even a call for national algorithm safety board (Macaulay 2017). Search results vary based on someone's gender; facial recognition works for some races and not others; curated news is more liberal. The headlines correctly warn against the hidden and unchecked biases of algorithms used in advertising, hiring, lending, risk assessment, etc. Hidden behind the apron of these headlines lies a tension between the idea that algorithms are neutral and organic when "the reality is a far messier mix of technical and human curating" (Dwork and Mulligan 2013, p. 35).

This false tension—algorithms as objective, neutral blank slates versus deterministic, autonomous agents—has implications for whether and how firms are responsible for the algorithms they develop, sell, and use. For example, algorithms-as-a-blank-slate would suggest minimal responsibility for the developers who craft the algorithm and suggests a *caveat user* approach to algorithmic accountability. Alternatively, the algorithm-as-autonomous-agent narrative (e.g., a black box (Pasquale 2015)) suggests the users have no say or accountability in how algorithms make decisions.

The current conversation about algorithms absolves firms of responsibility for the development or use of algorithms. Developers argue that their algorithms are neutral and thrust into fallible contexts of biased data and improper use by society. Users claim algorithms are difficult to identify let alone understand, therefore excluding users of any culpability for the ethical implications in use. Further, algorithms are so complicated and difficult to explain—even called unpredictable and inscrutable (Barocas et al. 2013; Desai and Kroll 2017; Introna 2016; Ziewitz 2016)—that assigning responsibility to the developer or the user is deemed inefficient and even impossible.

This article identifies whether firms developing algorithms have a responsibility for algorithms when in use, what those firms are responsible for, and the normative grounding for that responsibility. The goal of this article is to argue how firms that develop algorithms are responsible for the ethical implications of algorithms in use. I first conceptualize algorithms as value-laden in that algorithms create moral consequences, reinforce or undercut ethical principles, and enable or diminish stakeholder rights and dignity. For many within technology studies, law, and policy, this premise is not new (Akrich 1992; Bijker 1995; Friedman and Nissenbaum 1996; Johnson 2004; Latour 1992; Winner 1980). I offer a framework as to what we mean by value-laden algorithms in the first section to counter the claim that algorithms are neutral.

Less discussed, and the focus of the second section, is how algorithms are also an important part of a larger decision and influence the delegation of roles and responsibilities within an ethical decision. In other words, in addition to the design of value-laden algorithms, developers make a moral choice as to the delegation of who-does-what between algorithms and individuals within the decision. In the third section, I ground the normative obligations of firms in that I argue firms are responsible for the ethical implications of algorithms used in decision making based on an obligation created when the firm willingly sells into the decision-making context and based on the unique knowledge and abilities of the firm designing and developing the algorithm.

This article has implications for both ethical decision making and corporate accountability research. First, once the ethical implications of algorithms are understood, the design and development of algorithms take on greater meaning. Here, the type of accountability associated with the algorithm is framed as constructed in design as a product of both the type of decision in use and how large a role individuals are permitted to have in the algorithmic decision. Second, the theory of algorithmic accountability offered here pushes the boundaries of how we hold firms

[1] While Northpointe did not provide an explanation as to the factors contributing to the parole decision, Rodríguez, through talking to other prisoners with different scores, realized he was denied parole due to the answer for question 19: "Does this person appear to have notable disciplinary issues?" By changing that score from a "Yes" to a "No", his score went from an 8 to a 1 (Wexler 2017).

[2] For example, "Every Six Hours, the NSA Gathers as Much Data as Is Stored in the Entire Library of Congress." http://www.popsci.com/technology/article/2011-05/every-six-hours-nsa-gathers-much-data-stored-entire-library-congress.

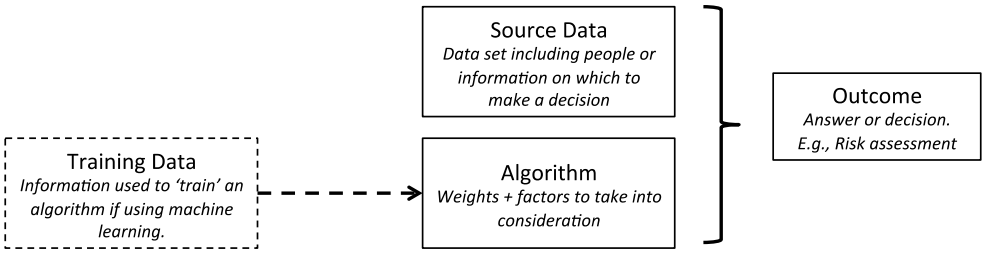

Fig. 1 Algorithm as producing "answer"

accountable for products that are working *as designed*. Previous work has focused on a type of product liability for when products or services go wrong (Brenkert 2000; Epstein 1973; Sollars 2003), yet the case of algorithms forces us to revisit examples of firms being responsible for when a product or service works as designed and still has ethical implications.

Finally, computer scientists are in the midst of an argument as to how algorithms can and should be transparent in order to be governed—including more autonomous algorithms such as machine learning, artificial intelligence, and neural networks (Burrell 2016; Desai and Kroll 2017; Howard 2014; Kroll et al. 2017; Ziewitz 2016; Selbst and Barocas 2018). Previous work has maintained that transparency is a precursor, perhaps an impossible precursor, to holding algorithms accountable. However, I address this dilemma by focusing on attributing accountability regardless of the level of algorithmic transparency designed. Firms can be held accountable for the ethical implications of the inscrutable algorithms they develop. "It's complicated" or "I do not know how it works" turns out to be an unsatisfying response to "who is responsible for this algorithm's value-laden biases?" Within business ethics, we attribute responsibility for many inscrutable and complicated decisions. I find creating inscrutable algorithms may, in fact, necessitate *greater* accountability afforded to the algorithm and the developer rather than less—counter to prevailing arguments within computer science, public policy, and law (Desai and Kroll 2017). We can hold firms responsible for an algorithm's acts even when the firm claims the algorithm is complicated and difficult to understand.

The article is organized as follows: I first use the case of risk assessment algorithms used in criminal justice decisions (e.g., sentencing) as illustrative of value-laden-ness of algorithms. This illustrative case also captures a particular use of algorithms in distributing social goods and the recognition of rights normally reserved for the state. I then leverage STS scholars Latour and Akrich to explain how these value-laden algorithms are not only biased but are designed to take on a role and associated responsibility within decision making and influence what individuals can do in an algorithmic decision. Finally, I justify why and under what conditions firms who develop algorithms should be held responsible for their ethical implications in use.

Ethical Implications of Algorithms

I turn to understand the outcome of concern or the object of responsibility: what is someone responsible *for* when it comes to algorithmic decisions? A persistent theme focuses on algorithms as blank slates mirroring back to society what is most accurate or efficient; the narrative of neutral algorithms would suggest firms have little to be responsible *for*. Figure 1 illustrates how algorithms are combined with a data set to produce an "answer" as currently understood in practice. As perhaps best defined by the most cited textbook on algorithms, an algorithm is a sequence of computational steps that transform inputs into outputs—similar to a recipe (Cormen 2009). Algorithms are viewed as maximizing efficiency or accuracy; computer scientists are, therefore, responsible for ensuring efficiency and accuracy (Seaver 2017).

In fact, algorithms are implemented with the hope of being more neutral (e.g., Barry-Jester et al. 2015), thereby suggesting that the decisions are better than those performed solely by individuals. By removing individuals from decisions—decisions such as sentencing, university admissions, prioritization of news—algorithmic decisions are framed as less biased without the perceived irrationality, discrimination, or frailties of humans in the decision. Within the narrative of neutrality, arguments acknowledging a biased algorithmic decision emphasize that the bias is due to the many ways individuals remain involved in the algorithmic decisions (Bozdag 2013).

One attraction of arguing that algorithms are neutral is the ability to avoid any form of technological determinism: in attributing values or biases to algorithms, scholars are concerned we would also attribute control to technology and thereby remove the ability of society to influence technology. Even further, identifying the value-laden-ness of algorithms could lead to a form of worship, where an algorithm's preferences are deemed unassailable and humans are left subservient to the whims of the algorithm (Desai and Kroll 2017).

In effect, the authors who argue this are conflating two ideas: whether or not a technology is value-laden and who controls the technology. Martin and Freeman argue these two mechanisms are independent and see technology as simultaneously value-laden yet under social control (Martin and Freeman 2004), where one need not claim technology as neutral to maintain control over it. Similarly, and focused on algorithms, Mittelstadt et al. note that algorithms are value-laden with biases that are "specified by developers and configured by users with desired outcomes in mind that privilege some values and interests over others"(2016). In other words, in creating the algorithm, developers are taking a stand on ethical issues and "expressing a view on how things ought to be or not to be, or what is good or bad, or desirable or undesirable" (Kraemer et al. 2011, p. 252).[3]

Below I use Northpointe's COMPAS algorithm in sentencing, as referenced in the introductory vignette, to illustrate how algorithms are not neutral but value-laden in that they (1) create moral consequences, (2) reinforce or undercut ethical principles, or (3) enable or diminish stakeholder rights and dignity.

Creating Moral Consequences

Critiques of risk assessment or sentencing algorithms have focused on whether the outcome of the algorithm is biased and harms particular groups of individuals (Angwin et al. 2016a; Skeem and Lowenkamp 2015). ProPublica, a nonprofit newsroom that produces investigative journalism, found that the COMPAS score proved remarkably unreliable in forecasting violent crime: only 20% of the people predicted to commit violent crimes actually went on to do so (Angwin et al. 2016a). More problematic, the investigative reporters also identified significant racial disparities: the algorithm wrongly labeled defendants as "future criminals" when they did not commit a crime at twice the rate for black defendants as white defendants (Angwin et al. 2016a). Further, white defendants were mislabeled as low risk, when they were not, more often than black defendants (Angwin et al. 2016a). Table 1 summarizes their findings.

COMPAS is a prime example of disparate impact by an algorithm (Barocas and Selbst 2016): where one group receives differential outcome outside the implicit norms of allocation (Colquitt 2001; Feldman et al. 2015).[4]

Table 1 Prediction fails differently for black defendants (Angwin et al. 2016a)

	White defendants (%)	Black defendants (%)
Labeled higher risk, but didn't re-offend	23.5	44.9
Labeled lower risk, yet did re-offend	47.7	28.0

In the sentencing case, the algorithms not only disproportionately impact a group of individuals, but the inequality (a higher sentence) also increases the likelihood the defendant will have lasting negative impact on life post-incarceration. Inequalities can exist, in other words, so long as they do not further harm the least advantaged in society (Rawls 2009). Putting low-risk offenders in prison with high-risk prisoners increases the likelihood they will re-offend (Andrews and Bonta 2010; Barry-Jester et al. 2015). The group disadvantaged—black defendants—are also the least fortune in the criminal justice system facing disproportionate "stop-and-frisk" incidents, car stops, arrests, and higher sentences, all else being equal (Gettman et al. 2016; Urbina 2013).[5]

Reinforcing or Undercutting Ethical Principles

Separate from the consequences of an algorithmic decision, an algorithm can either reinforce or violate ethical principles of the decision context. Algorithms rely upon a set of features—the attributes of the data set deemed important to the decision—as input. Which features of the data set are selected as important may be either appropriate or inappropriate for the decision at hand. Attorney General Eric Holder perhaps best summarizes this concern in regard to risk assessment algorithms in criminal justice:

> I am concerned that they [algorithms used in sentencing] inadvertently undermine our efforts to ensure individualized and equal justice, … Criminal sentences

[3] Similarly, algorithms act like design-based regulation (Yeung 2017) where algorithms can be used for the consistent application of legal and regulatory regimes (Thornton 2016, p. 1826); algorithms can enforce morality (Diakopoulos 2013)—while still being designed and used by individuals.

[4] For algorithms, in addition to directly coding the algorithm to prioritize one group more than any others, two mechanisms can also indirectly drive bias in the process: proxies and machine learning. Both are explored in the implications for practice and the ethics of design.

[5] Cathy O'Neil refers to these types of exacerbating impacts, where the algorithm (a) is developed to create systematic bias, (b) that impacts the less fortunate, and (c) does so with the volume and velocity attributed to big data initiatives, as weapons of math destruction (WMD) (O'Neil 2016). An algorithm can perpetuate injustices with *increasing* frequency. The technology appears to learn from current biases, create "answers" that are unjustly biased, and contributes to a new data set that is unjust upon which future algorithms will learn, thus creating a biased cycle of discrimination with little intervention required. The unjust bias feeds on itself.

> must be based on the facts, the law, the actual crimes committed, the circumstances surrounding each individual case, and the defendant's history of criminal conduct. They should *not* be based on unchangeable factors that a person cannot control, or on the possibility of a future crime that has not taken place (Holder 2014).

According to Holder, the ethical principles of the US criminal justice system dictate appropriate factors to consider in sentencing; the COMPAS algorithm violates those principles by design, because COMPAS utilizes unchangeable factors that a person cannot control such as a parent's criminal record or the first time someone was stopped by the police. Similarly, when an auto insurance algorithm is designed to consider your credit score as a more significant factor that your history of a DUI (O'Neil 2016), we should question whether the appropriate factors are used to judge the individual. In the case of the risk assessment algorithm, some factors are included (parental criminal history) while others are ignored (drug rehabilitation), which are incompatible with ethical principles of the decision.

Enabling and Diminishing Stakeholder Rights and Dignity

In addition to having adverse consequences or not following ethical principles, algorithms can be designed to undercut individuals' rights and dignity. Risk assessment algorithms such as COMPAS are kept secret, and defendants are not able to question the process by which their score was calculated. In a non-algorithmic sentencing, a probation officer may file a report, including a risk assessment of the defendant, and the prosecutor and defense attorney also make their case in court as to the appropriate sentence. The judge is able to query the individuals about the factors they each take into consideration. For risk assessment algorithms, the existence of the algorithm, the factors considered, and the weight given to each are kept secret by claiming the algorithm is proprietary (Smith 2016; Wexler 2017).[6]

Danielle Citron refers to this issue as technological due process (Citron 2007), arguing that "[t]his century's automated decision making systems combine individual adjudications with rulemaking while adhering to the procedural safeguards of neither." Algorithms are used across a range of what justice scholars would call the distribution of social goods such as education, employment, police protection, medical care, etc. Algorithms are used in decisions to terminate individuals' Medicaid, food stamps, and other welfare benefits as well as the "adjudication of important individual rights" (Citron 2007, p. 1253). More recently, algorithms have been used to categorize individuals as terrorists in creating the No Fly list (Hu 2016). As such, algorithms can constitute threats to due process rights and "deprive individuals of their liberty and property, triggering the safeguards of the Due Process Clauses of the Fifth and Fourteenth Amendments" (Citron 2007, p. 1281).

This need not be the state using an algorithm to diminish rights or dignity; private firms use algorithms to target teens online in a vulnerable state such as those who are depressed and anxious or when they feel insecure, worthless, defeated, and stressed (Garcia-Martinez 2017). Further, companies using algorithms to nudge consumers in a preferred direction can undercut the autonomy of decision makers (Helbing et al. 2017). Ryan Calo refers to the use of algorithms to define product searches based on consumers' hidden preferences as digital market manipulation (Calo 2014): consumers' autonomy could be undercut if their unrevealed, and perhaps even unknown, preferences are used against them in the market.

Reframing "Neutral" Algorithm

Rather than being neutral, if algorithms are value-laden with preferences for certain outcomes while still constructed by individuals in design, implementation, and use, then we have an open question if developers have a responsibility for algorithms in use, what firms are responsible for, and the normative grounding for that responsibility. Algorithms, and technology generally, are biased and designed for a *preferred* set of actions. Figure 2 reframes algorithmic decision making to include the ethical implications rendering the choice of factors, sourcing of the data, and assessment of the output more explicitly value-laden.

Algorithms as Value-Laden Actors Within Decisions

The ethical implications of algorithms outlined above are important to acknowledge not only because we should ensure biases are just and appropriate to the norms of the decision context, but also, as I turn to next, because value-laden algorithms become an important actor of a larger decision—an actor that determines the roles and responsibilities

[6] As noted by Northpointe's general manager, "The key to our product is the algorithms, and they're proprietary…We've created them, and we don't release them because it's certainly a core piece of our business. It's not about looking at the algorithms. It's about looking at the outcomes" (Smith 2016). It is difficult to fathom the human-centric version of such a stance: the probation officer, who may have been very good at predicting risk and gave "accurate" sentencing guidelines to the court, would state that she could not provide any explanation as to how she makes her judgments or what she takes into consideration. She would argue that doing so could jeopardize her job since she could then be replaced.

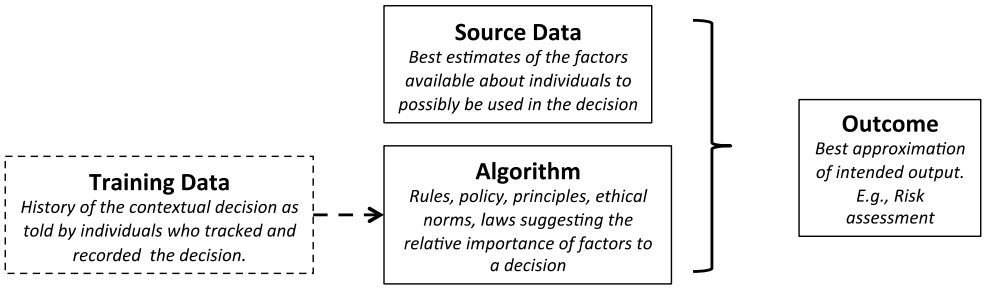

Fig. 2 Reframing data and algorithms as constructed—with biases throughout

of individuals in the decision. To claim that technology takes on roles and responsibilities within a system of actors is not without controversy. Algorithms have been referred to as actants (Tufekci 2015) as has technology more generally (Johnson and Noorman 2014), where material artifacts are designed to act within a system of material and non-material (i.e., human) actors that seeks to achieve a goal. Below, I leverage two scholars—Madeleine Akrich and Bruno Latour—to frame how algorithms impact the role and responsibilities of individuals and algorithms within a decision.

Role of Algorithms in Decisions

According to socio-technical studies (STS) scholar Madeleine Akrich, the design of technology is a projection of how the technology will work within a network of material and non-material actors. A car is designed with assumptions about the type of driver, how the roads are constructed, the number of passengers and how they will behave, the size of other cars on the road, etc. Cars have particular size openings (doors) and are designed at a width and height to both fit within the roads and keep individuals safe from other cars. While a plane may require a copilot, cars do not make such an assumption about what passengers will do. The safety of the passengers is designed into the car with airbags, seat belts, antilock brakes, collapsible front-ends, etc. as well as how the individuals and technologies will work together. As Akrich notes, "…A large part of the work of innovators is that of '*inscribing*' this vision of (or prediction about) the world in the technical content of the new object. I will call the end product of this work a 'script' or a 'scenario'" (Akrich 1992, p. 208). Designers of technological artifacts make assumptions about what the world will do and, relatedly, inscribe how their technology will fit into that world.

In terms of algorithms, Akrich's "script" is actually *less* obscure since the design is embodied in code that resembles language. Where the script behind a car or iPhone or toaster may require some imagination as to what the designer is saying, the algorithm comes in a form familiar to many—some even with comments throughout to explain the design. Algorithms are designed with assumptions about what is important, the type of data that will be available, how clean the data will be, the role of the actor imputing the data, and who will use the output and for what purpose. The sentencing algorithm assumes the data are in a certain form and, in effect, states that those data required for the algorithm to make a decision are most important.

Technologies as scripts survive outside the hands of the designer. Scripts are durable, and the technology's script becomes independent of the innovator once in use. Akrich uses the example of the two-handled Angolan hoe as made for women carrying children on their backs (Akrich 1992, p. 208). The hoe exists with this biased script—giving preference to women carrying children—decades later. In the sentencing algorithm case above, the factors taken into consideration, such as COMPAS algorithm's 137 questions, exist after the algorithm is put into use. Changing the hoe's, the algorithm's, or a car's design after production is difficult. Importantly, while technology and algorithms are constructed by humans, technology's scripts endure to influence the behavior, acts, and beliefs of individuals.[7]

These technologies survive to have biases that are value-laden (Friedman and Nissenbaum 1996; Johnson 2004) or have politics (Winner 1980). The design of the car to fit within roads and survive most crashes is value-laden: the script acknowledges the validity of the current road design and preferences certain types of people (by size, weight, gender) to survive a crash.[8] The quintessential example within technology studies is Langdon Winner's analysis of bridges

[7] Latour, Akrich, this article, and others (Martin and Freeman 2004) remain outside the technological determinism versus social constructivism divide. As Akrich notes: "technological determinism pays no attention to what is brought together, and ultimately replaced, by the structural effects of a network. By contrast social constructivism denies the obduracy of objects and assumes that only people can have the status of actors" (p. 206). Martin and Freeman rightly separate the idea of technology's value-laden-ness and social control as independent attributes: a technology can have a value-laden bias while also being influenced by society in general and by individuals.

[8] A recent example concerning crash tests and female crash-test dummies confirms this longstanding issue (Shaver 2012). Cars were only designed and tested for the safety of men until 2011.

on the road to Jones Beach. These bridges were designed at a height that would preclude public buses (and people who took public buses) from accessing Jones Beach, thus prioritizing those with cars and excluding those who rely on public transportation. The technology's script answers who matters, which group is important, who counts, which race/ethnicity is included and delineated. In the sentencing example, the algorithm states that a defendant's paternal criminal history is important but the defendant's own recovery from addiction is not. The algorithm-as-script makes assumptions as to the accuracy of the data and how the output will be used. Akrich suggests the following thought experiment which is of particular importance for algorithms:

> …How can the prescriptions encoded in the mechanism be brought out in words? By replacing them by strings of sentences (often in the imperative) that are uttered (silently and continuously) by the mechanisms for the benefit of those who are mechanized: do this, do that, behave this way, don't go that way, you may do so, be allowed to go there. Such sentences look very much like a programming language.

Algorithm's Delegation of Roles and Responsibilities in Decisions

Technologies, such as algorithms, influence a group of actors assembled to perform a task. Algorithmic biases not only impact the achievement of the task as well as whether and how ethical norms are respected, but also the function and role of the other actors in the decision. Latour uses the combination of a door and a door groomer to illustrate how tasks may be delegated between material and non-material actors. The door hinge allows us to gain access to a room without tearing down walls and rebuilding them.[9] The combination of the door, the hinge, and the doorman creates the opportunity to walk through a wall without leaving a gaping hole in the wall. Similarly, a system of airbags, seat belt, driver, and an annoying chime combine to secure the driver in the event of a crash (Latour 1992). In the case of the sentencing algorithm, COMPAS works within a system of actors in the court to adjudicate the sentence including the judge, probation officer, defense attorney, defendant, prosecutor, clerks, etc.

At a minimum, technologies alleviate the need for others to do a task. In the case of Latour's seat belt, making the seat belt automatic—attaching the seat belt to the door so that it is in place automatically—alleviates the driver from the responsibility to ensure the seat belt is used. In the case of doors, hydraulic door hinges ensure the door is closed gently without the need of a human door groomer. In the case of sentencing algorithms, COMPAS makes sense of the defendants' profile and predicts their risk assessment, thereby alleviating the need of the probation officer or judge from making that judgment. As Latour rightly summarizes, "every time you want to know what a nonhuman does, simply imagine what other humans or other nonhumans would have to do were this character not present" (p. 155). This delegation of tasks is a choice, and this delegation is constructed and constantly up for deliberation. The divvying up of tasks between material and non-material actors (i.e., algorithms and individuals) within a safety system, sentencing system, or go-through-the-wall system appears as a *fait accompli* when the system works. However, this delegation as to who-does-what deserves careful consideration.

Importantly, the substitution of technology for human is not a perfect substitution: as Latour notes, "springs do the job of replacing grooms, but they play the role of a very rude, uneducated, and dumb porter who obviously prefers the wall version of the door to its hole version. They simply slam the door shut" (p. 157). Also, due to their prescriptions, these door springs have biases and "*discriminate* against very little and very old persons" (p. 159, italics in original). Sentencing algorithms in sentencing illustrate a similar problem with unjust biases perpetuating human discrimination. Similarly, an algorithm for university admittance could be as discriminatory by design or the algorithm could be trained on data with historical biases.[10] Replacing the discriminatory human with a biased technology does not erase the discrimination.

Technologies, such as algorithms, are designed to perform a task with a particular moral delegation in mind. This moral delegation by designers impacts the moral behavior of other actors. In the case of the doors, designers decide "either to discipline the people or to substitute for the unreliable humans a delegated nonhuman character whose only function is to open and close the door. This is called a door-closer or a groom" (Latour 1992, p. 157). The hydraulic door groom takes on the responsibility to close the door.

Here, I suggest that computer scientists perform the same delegation of tasks in designing an algorithm. Just as there is a distribution of competences between technology and individuals, there is also a distribution of associated responsibility. Latour suggests thinking about the morality in extreme cases: where the design of the car stipulates that the seat belt must

[9] As Latour notes, "we have delegated….to the hinge the work of reversibly solving the wall-hole dilemma" (Latour 1992, p. 155).

[10] In this way, Latour notes that technology—including algorithms—is anthropomorphic: "first, it has been made by humans; second, it substitutes for the actions of people and is a delegate that permanently occupies the position of a human; and third, it shapes human action by prescribing back" what humans should do (p. 160).

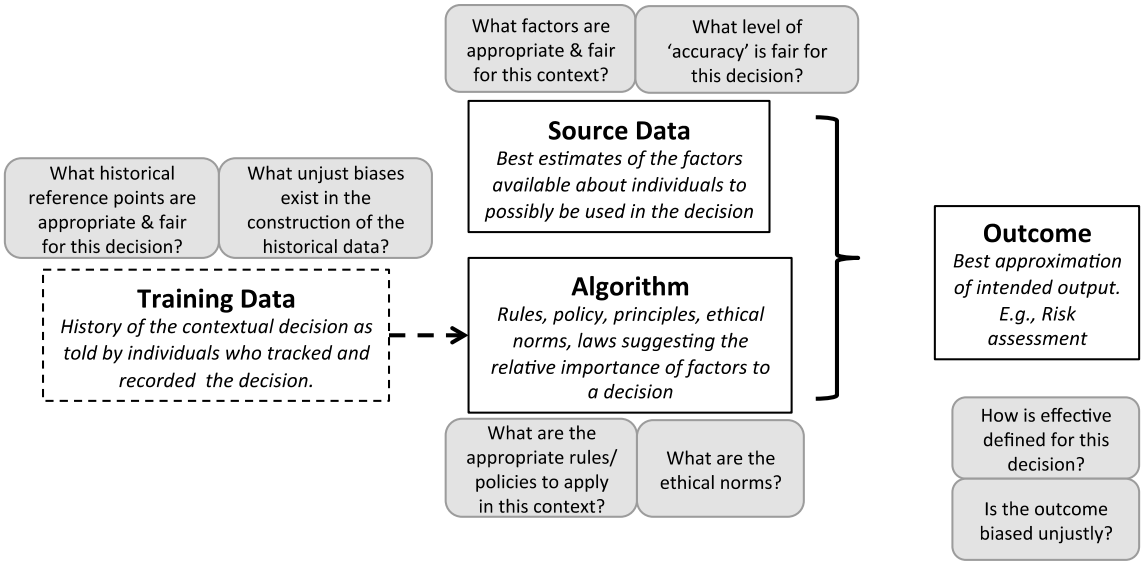

Fig. 3 Adding in missing masses to algorithm decision-making process

be fastened before the car could start versus where the car is designed without any nudges for the driver.

> Worse yet – the design where 'a seat belt that politely makes way for me when I open the door and then straps me as politely but very tightly when I close the door'… The program of action "IF a car is moving, THEN the driver has a seat belt" is enforced…I cannot be bad anymore. I, plus the car, plus the dozens of patented engineers, plus the police are making me be moral (p. 152).

In delegating the task of driver safety to the technology, the designer alleviates the individual from having to take on that responsibility.

Delegating a task to a technology—such as a seat belt or an algorithm—does not remove the associated responsibility for that task. Latour uses physicists looking for "missing mass" in the universe as a metaphor for sociologists or ethicists looking for missing responsibility in a system of technologies and individuals. Latour suggests we start looking in material actors for the missing masses "who make up our morality" (Latour 1992, pp. 152–153). Figure 3 makes explicit (some of) the missing masses in algorithmic decision making. By adding back the questions, we are silently asking and perhaps delegating to algorithms in design, Latour's missing masses crowd out the role of the algorithm in Fig. 3.

Designing an Algorithm Prescribes the Delegation of Responsibilities in Decisions

This delegation of roles and responsibilities of the decision and the value-laden-ness of algorithms are important ethical decisions we continually make in design and development—whether firms acknowledge the decisions or not. Each box in Fig. 3 can be answered by an algorithm or a human, and designers decide the delegation of roles and responsibilities between humans and algorithms when creating an algorithm. This decision of how roles and responsibilities are allocated to human and algorithm is performed by the engineer. For Latour, "It is the complete chain that makes up the missing masses, not either of its extremities. The paradox of technology is that it is thought to be at one of the extremes, whereas it is the ability of the engineer to travel easily along the whole gradient and substitute one type of delegation for another that is inherent to the job" (1992, p. 166).

Ignoring the moral delegation of roles, responsibilities, and the missing masses does not make them disappear or become less important. As noted by Richard Rudner in regard to the value-laden decision throughout the scientific process, "To refuse to pay attention to the value decisions which must be made, to make them intuitively, unconsciously, haphazardly, is to leave an essential aspect of scientific method scientifically out of control" (1953, p. 6). The decisions about biases, roles, and responsibilities should be brought into the foreground for designers as in Fig. 3. When algorithmic decision making is anemically framed as in Fig. 1, Latour's 'masses that make up our morality' go missing, and the delegation of responsibility appears to be inevitable and taken-for-granted.[11] No one is accountable

[11] As Akrich notes, "two vital questions start to come into focus. The first has to do with the extent to which the composition of a technical object constrains actants in the way they relate to both the object and to one another. The second concerns the character of these actants and their links, the extent to which they are able to reshape the

for the decision as to who can and should answer the questions in Fig. 3. However, the argument here is that the moral delegation of roles and responsibilities is still occurring in the scripts of the algorithm as inscribed in design.

In other words, in addition to the design of value-laden biases, firms make a moral choice as to the delegation of tasks and responsibilities between algorithms and individuals in design. This choice, if ignored, will not only be out of control as noted by Rudner, but the construction of biases and the delegations of roles, responsibilities, and missing masses will continue unquestioned.

Accountability for Algorithmic Decision Making

In this article, I have conceptualized how algorithms are value-laden rather than neutral, where algorithms are inscribed with a preferred set of outcomes with ethical implications. The value-laden biases are important to acknowledge not only because we should ensure algorithms are just, conform to principles and norms of the decision, and enable rather than diminish rights ("Ethical Implications of Algorithms" section), but also because algorithms are an important part of a larger decision and influence the delegation of roles and responsibilities within ethical decisions ("Algorithms as Value-Laden Actors Within Decisions" section). I now turn to explore why firms have a unique obligation in the development of algorithms around the ethical implications and roles of an algorithm in an ethical decision.

Accountability and Inscrutable Algorithms

Previous approaches to algorithmic accountability amount to a dichotomous choice. At one extreme, algorithms are value-neutral and determined by their use, with accountability falling exclusively on the users or even "society" (Kraemer et al. 2011). At the other end of the spectrum is a more deterministic argument, whereby algorithms are controlling yet obscure, powerful yet inscrutable (Neyland 2016; Ziewitz 2016) and veer toward algorithms as beyond our control and the primary actors. For example, Desai and Kroll (2017) argue

> Some may believe algorithms should be constructed to provide moral guidance or enforce a given morality. Others claim that moral choices are vested with a system's users and that the system itself should be neutral, allowing all types of use and with moral valences originating with the user. In either case, … the author's deference to algorithms is a type of "worship" that reverses the skepticism of the Enlightenment. Asking algorithms "to enforce morality" is not only a type of idolatry, it also presumes we know whose morality they enforce and can define what moral outcomes are sought. [Underlining added].

Desai and Kroll rightly identify the challenge we face in identifying the moral norms an algorithm either supports or undercuts. However, algorithms are currently enforcing morality by preferencing outcomes and the roles of others in the decision, whether or not we acknowledge that enforcement and seek to govern the design decisions. The question is, who is responsible for the ethical implications rather than whether or not the algorithm provides moral guidance.

When developers design the algorithm to be used in a decision, they also design how accountability is delegated within the decision.[12] Sometimes algorithms *are designed* to absorb the work and associated responsibility of the individuals in the decision by precluding users from taking on roles and responsibilities within the decision system—e.g., inscrutable algorithms designed to be more autonomous and with less human intervention (Barocas et al. 2013; Desai and Kroll 2017; Introna 2016; Ziewitz 2016). For example, the COMPAS algorithm was designed to preclude individuals from understanding how it works or from taking any responsibility for how it is implemented. Importantly, this is a design choice because other risk assessment algorithms are designed to be more open, thereby delegating more responsibility for the decision to individuals (Kramer 2017).

Importantly, firms can be held accountable for inscrutable systems. Inscrutable algorithms that are designed to minimize the role of individuals in the decision take on more accountability for the decision. In fact, one should be suspect of the inscrutable defense: when systems have been called inscrutable in order to avoid being effectively governed such as Enron's accounting, banks' credit-default swaps, or a teenager's reasons behind a bad grade. The inscrutable defense ("It's too complicated to explain") does not absolve a firm from responsibility; otherwise, firms would have an incentive to create complicated systems to avoid accountability. Firms and individuals are held accountable for decisions and products that are difficult to explain. Some cars are designed to be maintained by anyone including the owner;

Footnote 11 (continued)

object, and the various ways in which the object may be used. Once considered in this way, the boundary between the inside and the outside of an object comes to be seen as a *consequence* of such interaction rather than something that determines it" (Akrich 1992).

[12] Interesting challenges arise for algorithms with learning capacities, as they defy the traditional conception of designer responsibility--programmers see themselves as less involved in the final product since the algorithm "learns" from the data rather than being 100% coded directly by the programmer. See also Mittelstadt et al. (2016).

others are designed to require a professional license where the manufacturer takes on responsibility to ensure the car is working properly. Importantly, firms develop products *knowing* they are going to be held accountable.

According to the argument herein, inscrutable algorithms—designed to be difficult to understand and argued to be hard to explain—may force greater accountability on the designer to own the algorithmic decision since their design of the algorithm has precluded anyone else from taking on a larger role in the decision when in use. Previous arguments against algorithmic transparency have centered on pitting fairness against accuracy or as being inefficient or just difficult to accomplish (Ananny and Crawford 2016; Jones 2017; Kroll et al. 2017). Creating inscrutable algorithms precludes users from taking responsibility for the ethical implications identified above and places the responsibility of the ethical implications on the firm who developed the algorithm. The design of the algorithm not only scripts what users can do but also the reasonable expectations of users to take responsibility for the use of the algorithm.

Why Firms are Responsible for the Algorithms they Develop

Within the arguments of this article, the onus now shifts to the developer of the algorithm to take responsibility for not only the ethical implications of the algorithm in use but also how roles will be delegated in making a decision. Alternatively, developers can design the algorithm to allow users to take responsibility for algorithmic decisions. However, the responsibility for such design decisions is on the knowledgeable and uniquely positioned developers. This obligation is based on two arguments. First, a firm's obligation for the ethical implications of an algorithm is created because the firm is knowledgeable as to the design decisions and is in a unique position to inscribe the algorithm with the value-laden biases as well as roles and responsibilities of the algorithmic decision. Developers are those most capable of enacting change in the design and are sometimes the *only* individuals in a position to change the algorithm. In other words, by willingly creating an algorithm that works in a value-laden and particular manner, firms voluntarily become a party to the decision system and take on the responsibility of the decision to include the harms created, principles violated, and rights diminished by the decision system. How much responsibility and for what acts depends on how the algorithm is designed. In fact, as is argued here, the more the algorithm is constructed as inscrutable and autonomous, the *more* accountability attributed to the algorithm and the firm that designed the algorithm.

Second, an obligation is created when the firm developing the algorithm willingly enters into the decision context by selling the algorithm for a specific purpose. Selling an algorithm to the courts to be a risk assessment tool creates an obligation for the firm as a member of the criminal justice community. In social contract terms, firms that develop algorithms are members of the community to which they sell the algorithm—e.g., criminal justice, medicine, education, human resources, military, etc.—and create an obligation to respect the norms of the community as a member (Donaldson and Dunfee 1994). If a company does not wish to abide by the norms of the decision (e.g., being transparent for due process rights of defendants) or be accountable for the moral consequences and rights impacted by a pivotal decision in society, then the firm should not be in that business and not sell the algorithm into that particular context. By entering the market, the firm voluntarily takes on the rules of that market including the norms of the decisions it is facilitating.

For example, the decision to manufacture drones for the military created an obligation for defense contractors to understand the rules of engagement for our military using the drones. For a company developing manufacturing equipment, the designer must understand how the plant worker can be expected to work given not only the laws governing safety but also the norms of the industry. (This is normally called human factors engineering.) Algorithms are no different: when companies decide to develop and sell algorithms within a decision context, the organization willingly takes on the obligation to understand the values of the decision to ensure the algorithms' ethical implications is congruent with the context.

Discussion and Conclusion

This article has shown how algorithms act as structuring agents in both mundane and key decisions and developed how and why firms are responsible for the design of algorithms within a decision. First, I offered a systematic account of the value-laden-ness of algorithms. Second, I leveraged STS scholars Latour and Akrich to frame algorithms as actors in ethical decision making—delegated tasks and responsibilities akin to other actors in the decision. Third, I grounded the normative obligation of developers for the ethical implications of algorithms. If a firm's technology, such as an algorithm, acts to influence others, then companies could be held accountable for the acts, biases, and influence of their technology. I conclude with the implications for corporate responsibility, fiduciary duties, transparency, and research on algorithms.

Corporate Responsibility for Algorithms

Based on the arguments here, responsibility for algorithmic decision making is constructed in the design and development of the algorithm. Yet, corporate responsibility about

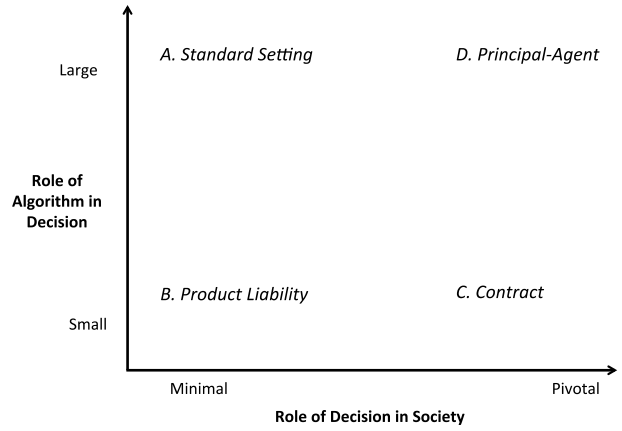

Fig. 4 Firm responsibility for algorithms

products and services centers on situations where something goes wrong: a breach of a contract, product liability, or a tort harm created by a company. And, business ethics struggles to identify how and when firms are responsible for the use of a product that is working correctly and as designed (Brenkert 2000; Sollars 2003). A parallel argument about gun manufactures, where the correct use of the product can cause harm, has focused on marketing and distribution (Byrne 2007). Brenkert goes further to include not only product defects, but general harms caused by products as with gun manufacturers, "in which a non-defective product does what it is designed to do, but, because of the social circumstances in which it comes to be used, imposes significant harms through the actions of those who are using it in ways in which it was not supposed to be used" (Brenkert 2000, p. 23). Algorithmic harms can differ as the unjust biases can be due to the "correct" use.

One possible avenue for future corporate responsibility research is linking the role of the algorithm in a decision with the responsibility of the firm as shown in Fig. 4. In other words, firms (1) construct algorithms to take on a large or small role in a decision (y-axis) and (2) sell that algorithm to be used within a specific context (x-axis); both decisions contribute to the appropriate type of responsibility we expect of the firm. For example, a firm that develops an algorithm to take on a larger role in a decision of minimal societal importance—e.g., deciding where to place an ad online—could be seen as standard setting as to appropriate biases as well as the delegation of roles and responsibility encoded in the design. The firm acts as an expert in heavily influencing the decision including what factors are important and appropriate for a decision. Alternatively, if the role of the algorithm in a decision is minimized, by providing tools to allow users to revisit how the algorithm works, the firm would have more of a traditional handoff of a product with associated (minimal) responsibility around product liability. The difference between A and B in Fig. 4 would be the role of individuals using the algorithm as inscribed in the algorithm design; greater agency of the individual over the algorithm in use means less accountability attributed to the algorithm within the decision.

For decisions seen as pivotal in the life of individuals (O'Neil 2016)—whereby the decision provides a gatekeeping function to important social goods such as sentencing, allocation of medical care, access to education, etc.—the expected relationship could be akin to a principle–agent relationship where the algorithm acts as an agent for the design firm. The developer scripts the agent (algorithm) and the algorithm carries out its prescribed duties (e.g., Johnson and Noorman (2014); Powers and Johnson). Delegating decisions to drones in military situations takes on similar scrutiny where the developer (a contractor for the government or the military itself) remains responsible for the actions of the agent. If the developer wishes the algorithm to take a smaller role in a pivotal decision, the responsibility may be closer to a contract with a responsibility to remain engaged for the duration of the algorithm's use in case the role changes because the decision is pivotal. Key for future work about appropriate corporate responsibility would be acknowledging that how the firm designed the algorithm to take on a role within the decision implies an associated responsibility for the decision itself.

Ethics of Algorithmic Design

Positioning the algorithm as having an important role within the larger ethical decision highlights three areas of concern for designing and coding algorithms. First, developing accountable algorithms requires identifying the principles and norms of decision making, the features appropriate for use, and the dignity and rights at stake in the situated use of the algorithm. Algorithms should be designed understanding the delegation of roles and responsibilities of the decision system. Second, give the previous section, algorithms should be designed and implemented toward the appropriate level of accountability within the decision, thereby extending the existing work on algorithm accountability (Kroll et al. 2017).

Finally, the ethical implications of algorithms are not necessarily hard-coded in the design and firms developing algorithms would need to be mindful of indirect biases. For COMPAS, individuals across races do not have an equal chance of receiving a high-risk score. The question is, why? Assuming COMPAS did not design the algorithm to code "Black Defendant" as a higher risk directly, why are black defendants more likely to be falsely labeled as high risk when they are not? Algorithms can be developed with an explicit goal such as to evade detection of pollution

by regulators as with Volkswagon (LaFrance 2017).[13] For algorithms, two mechanisms can also indirectly drive bias in the process: proxies and machine learning. First, when a feature cannot or should not be used directly (e.g., race), an algorithm can be designed to use closely correlated data as a proxy that stands in for the non-captured feature. While race is not one of the questions for the risk assessment algorithms, the survey includes questions such as "Was one of your parents ever sent to jail or prison?" which can be highly correlated with race given drug laws and prosecutions in the 1970s and 1980s (Angwin et al. 2016a; Gettman et al. 2016; Urbina 2013). For example, researchers were able to identify individuals' ethnicity, sexual orientation, or political affiliation from the person's Facebook "likes" (Tufekci 2015). Similarly, loan terms or pricing should not vary based on race, but banks, insurance companies, and retail outlets can target based on neighborhoods or social connections, which can be highly correlated with race (Angwin et al. 2017; Waddell 2016).[14] In this case, basing scores on the father's arrest record or the neighborhood where the defendant lives or "the first time you were involved with the police" can prove to be a proxy for race (Andrews and Bonta 2010; Barry-Jester et al. 2015; O'Neil 2016).

In addition to using proxies, value-laden algorithms could also be due to training the algorithm on biased data with machine learning. Some algorithms learn which factors are important to achieving a particular goal through the systematic examination of historical data as shown in Fig. 1. The COMPAS algorithm is designed to take into consideration a set number of factors and weight each factor according to its relative importance to a risk assessment. A classic example used by Cynthia Dwork, a computer scientist, the Distinguished Researcher at Microsoft Research, and quoted at the beginning of this article, is of university admissions. In order to identify the best criteria by which to judge applicants, a university could use a machine learning algorithm with historical admissions, rejection, and graduation records going back decades to identify what factors are related to "success." Success could be defined as admittance or as graduating within 5 years or a particular GPA (or any other type of success). Importantly, historical biases in the training data will be learned by the algorithm, and past discrimination will be coded into the algorithm (Miller 2015). If one group—women, minorities, individuals of a particular religion—was systematically denied admissions or even underrepresented in the data, the algorithm will learn from the biased data set.

Biased training data are an issue that crosses contexts and decisions. Cameras trained to perform facial recognition often fail to correctly identify for certain races: a facial recognition program could recognize white faces but was less effective detecting faces of non-white races. The data scientist "eventually traced the error back to the source: In his original data set of about 5000 images, whites predominated" (Dwoskin 2015). The data scientist did not write the algorithm to focus on white individuals; however, the data he used to *train* the algorithm included predominately white faces. As noted by Aylin Caliskan, a postdoc at Princeton University, "AI is biased because it reflects effects about culture and the world and language…So whenever you train a model on historical human data, you will end up inviting whatever that data carries, which might be biases or stereotypes as well" (Chen 2017).

Machine learning biases are insidious because the bias is yet another level removed from the outcome and more difficult to identify. In addition, the idea behind machine learning—to use historical data to teach the algorithm what factors to take into consideration to achieve a particular goal—appears to further remove human bias, until we acknowledge that the historical data were created by biased individuals. Machine learning biases have the veneer of objectivity when the algorithm created by machine learning can be just as biased and unjust as one written by an individual.

Transparency

Calls for algorithmic transparency continue to grow: yet full transparency may be neither feasible nor desirable (Ghani 2016). Transparency as to how decisions are made can allow individuals to "game" the system. People could make themselves algorithmically recognizable and orient their data to be viewed favorably by the algorithm (Gillespie 2016), and gaming could be available to some groups more than others, thereby creating a new disparity to reconcile (Bambauer 2017). Gaming to avoid fraud detection or avoid SEC regulation is destructive and undercuts the purpose of the system. However, algorithmic opacity is also framed as a form of proprietary protection or corporate secrecy (also Pasquale 2015), where intentional obscurity is designed to avoid scrutiny (Burrell 2016; Diakopoulos 2015; Pasquale 2015).

Based on the model of algorithmic decision making in Fig. 4, calls for transparency in algorithmic decision making may need to be targeted for a specific purpose or type of decision. Annany and Crawford rightly question the quest for transparency as an overarching and unquestioned

[13] "They knew that during these tests, regulators would use specific parameters. So they wrote logic that—if those parameters were selected—the engine would run in a special mode," thereby masking the fact that the diesel engines actually produced up to 40×the federal limit (Larson 2017).

[14] "In some cases, insurers such as Allstate, Geico, and Liberty Mutual were charging premiums that were on average 30 percent higher in zip codes where most residents are minorities than in whiter neighborhoods with similar accident costs" (Angwin et al. 2017).

goal (Ananny and Crawford 2016). For example, the transparency to identify unjust biases may be different from the transparency for due process. Similarly, the transparency needed for corporate responsibility in the principal–agent relationship in Fig. 4 (a large role of the algorithm in a pivotal decision) would differ from the transparency needed for an algorithm that decides where to place an ad. Further, transparency can take on different forms. Techniques to understand the output based on changing the input (Dattam et al. 2016) may work for journalistic inquiries (Diakopoulos and Koliska 2017) but not for due process in the courts where a form of certification may be necessary.

Importantly, this range of transparency is possible. For example, a sentencing algorithm in Pennsylvania is being developed by a public agency, and the algorithms are open to the public for analysis (Smith 2016). Similarly, a company CivicScape released its algorithm and data online in order to allow experts to examine the algorithm for biases and provide (Wexler 2017). In fact, Wexler describes two competing risk assessment algorithms—one secret and one disclosed to defense attorneys—and both are competitive in the market. Based on the arguments here, the level and type of transparency would be a design decision and would need to adhere to the norms of the decision context. If a firm does not wish to be transparent about the algorithm, they need not be in a market focused on pivotal decisions allocating social goods with due process norms.

Implications for Ethical Decision-Making Theory

Just as ethical decision making offers lessons for algorithmic decisions, so to acknowledging the value-laden role of algorithms in decisions has implications for scholarship in decision making. First, more work is needed to understand how individuals make sense of the algorithm as contributing to the decision and the degree of perceived distributive and procedural fairness in an algorithmic decision. For example, Newman et al. (2016) empirically examine how algorithmic decisions within a firm are perceived as fair or unfair by employees. Recent work by Derek Bambauer seeks to understand the condition under which algorithmic decisions are accepted by consumers (Bambauer 2017).

Algorithms will also impact the ability of the human actors within the decision to make ethical decisions. Group decision making and the ability of individuals to identify ethical issues and contribute to a discussion could offer a road map as to how to research the impact of algorithms as members of a group decision (e.g., giving voice to values Arce and Gentile 2015). While augmented labor with robots is regularly examined, we must next consider the ethics and accountability of algorithmic decisions and how individuals are impacted by being a part of the algorithmic decision-making process with non-human actors in the decision.

Fiduciary Duties of Coders and Firms

The breadth and importance of the value-laden decisions of algorithms suggest greater scrutiny of designers and developers of algorithms used for pivotal decisions. If algorithms act as silent structuring agents deciding who has access to social goods and whose rights are respected, as is argued here, algorithmic decisions would need oversight akin to civil engineers building bridges, CPAs auditing firms, and lawyers representing clients in court. Similar to calls for Big Data review boards (Calo 2013), algorithms may need a certified professional involved for some decisions. Such professionalized or certified programmer would receive not only technical training but also courses on the ethical implication of algorithms. As noted by Martin (2015), many data analytics degrees do not fall under engineering schools and do not have required ethics courses or professional certification.

Research on Algorithms

Finally, firms should do more to support research on algorithms. Researchers and reporters run afoul of the CFAA, the Computer Fraud and Abuse Act, when performing simple tests to identify unjust biases in algorithms (Diakopoulos 2015; Diakopoulos and Koliska 2017). While the CFAA was designed to curtail unauthorized access to a protected computer, the act is now used to stop researchers from systematically testing output and service of websites based on different user types (Kim 2016). For example, researchers can violate the current version of the CFAA when changing a mock user profile to see whether Facebook's NewsFeed shows different results based on gender (Sandvig et al. 2016), whether AirBnB offers different options based on the race of the user, or to test whether Google search results are biased (Datta et al. 2015). And firms can make researchers' jobs harder even without the CFAA. After Sandvig et al. published their analysis on Facebook's NewsFeed, companies modified the algorithm to render the research technique used ineffective. Such tactics, whether using the CFAA or obscuring algorithms, serve to make researchers jobs harder in attempting to hold corporations accountable for their algorithmic decisions. Modifying the CFAA is one important mechanism to help researchers.

Conclusion

Algorithms impact whether and how individuals have access to social goods and rights, and how algorithms are developed and implemented within managerial decision making is critical for business ethics to understand and research.

We can hold firms responsible for an algorithm's acts even when the firm claims the algorithm is complicated and difficult to understand. Here, I argue, the deference afforded to algorithms and associated outsized responsibility for decisions constitutes a design problem to be addressed rather than a natural outcome of identifying the value-laden-ness of algorithms.

Acknowledgements The author is grateful for support from the National Science Foundation under Grant No. 1649415. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This paper was greatly improved based on feedback from Professors Ryan Calo, Gaston de los Reyes, and Bobby Parmar.

Compliance with Ethical Standards

Conflict of interest Kirsten Martin declares that he/she has no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

Akrich, M. (1992). The description of technological objects. In W. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 205–224). Cambridge, MA: MIT Press.

Ananny, M., & Crawford, K. (2016). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*. https://doi.org/10.1177/1461444816676645.

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy, and Law, 16*(1), 39.

Angwin, J., Larson, J., Kirchner, L., & Mattu, S. (2017). Minority neighborhoods pay higher car insurance premiums than white areas with the same risk. *ProPublica*. https://www.propublica.org/article/minority-neighborhoods-higher-car-insurance-premiums-white-areas-same-risk. Accessed 12 June 2017.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016a). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Accessed 3 Aug 2016.

Angwin, J., Parris Jr, T., & Mattu, S. (2016b). Breaking the black box: When algorithms decide what you pay. *ProPublica*. https://www.propublica.org/article/breaking-the-black-box-when-algorithms-decide-what-you-pay. Accessed 11 Oct 2016.

Arce, D. G., & Gentile, M. C. (2015). Giving voice to values as a leverage point in business ethics education. *Journal of Business Ethics, 131*(3), 535–542.

Bambauer, D. E. (2017). Uncrunched: Algorithms, Decision making, and Privacy, Second Annual Digital Information Policy Scholars Conference, George Mason University Antonin Scalia Law School, Arlington, VA. (Apr. 28, 2017).

Barocas, S., Hood, S., & Ziewitz, M. (2013). Governing algorithms: A provocation piece. http://dx.doi.org/10.2139/ssrn.2245322. Accessed 14 June 2017.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104,* 671.

Barry-Jester, A. M., Casselman, B., & Goldstein, D. (2015). *The new science of sentencing*. The Marshall Project. https://www.themarshallproject.org/2015/08/04/the-new-science-of-sentencing. Accessed 15 Aug 2016.

Bijker, W. (1995). *Of bicycles, bakelite, and bulbs: Towards a theory of sociological change*. Boston MA: MIT Press.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology, 15*(3), 209–227.

Brenkert, G. G. (2000). Social products liability: The case of the firearms manufacturers. *Business Ethics Quarterly, 10*(01), 21–32.

Brown, K. (2016). When Facebook decides who's a terrorist. *Fusion*. http://fusion.net/story/356354/facebook-kashmir-terrorism/. Accessed 11 Oct 2016.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 2053951715622512.

Byrne, E. F. (2007). Assessing arms makers' corporate social responsibility. *Journal of Business Ethics, 74*(3), 201–217.

Calo, R. (2013). Consumer subject review boards: A thought experiment. *Stanford Law Review Online, 66,* 97.

Calo, R. (2014). Digital market manipulation. *George Washington Law Review, 82*(4), 995.

Chen, A. (2017). AI picks up racial and gender biases when learning from what humans write. *The Verge*. https://www.theverge.com/2017/4/13/15287678/machine-learning-language-processing-artificial-intelligence-race-gender-bias. Accessed 12 June 2017.

Citron, D. K. (2007). Technological due process. *Washington University Law Review, 85,* 1249.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*(3), 386.

Cormen, T. H. (2009). *Introduction to algorithms*. Cambridge, MA: MIT press.

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies, 2015*(1), 92–112.

Desai, D. R., & Kroll, J. A. (2017). Trust but verify: A guide to algorithms and the law. *Harvard Journal of Law and Technology*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2959472.

Dewey, C. (2016). Facebook has repeatedly trended fake news since firing its human editors. *The Washington Post*. https://www.washingtonpost.com/news/the-intersect/wp/2016/10/12/facebook-has-repeatedly-trended-fake-news-since-firing-its-human-editors/?tid=sm_tw&utm_term=.b795225d264d. Accessed 2 Dec 2016.

Diakopoulos, N. (2013). Sex, violence, and autocomplete algorithms. *Slate*. http://www.slate.com/articles/technology/future_tense/2013/08/words_banned_from_bing_and_google_s_autocomplete_algorithms.html. Accessed 15 Aug 2016.

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism, 3*(3), 398–415.

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism, 5*(7), 809–828.

Donaldson, T., & Dunfee, T. W. (1994). Toward a unified conception of business ethics: Integrative social contracts theory. *Academy of Management Review, 19*(2), 252–284.

Dwork, C., & Mulligan, D. K. (2013). It's not privacy, and it's not fair. *Stanford Law Review Online, 66,* 35.

Dwoskin, E. (2015). How social bias creeps into web technology. *Wall Street Journal*. http://www.wsj.com/articles/computers-are-showing-their-biases-and-tech-firms-are-concerned-1440102894?mod=rss_Technology. Accessed 12 Aug 2016.

Epstein, R. A. (1973). A theory of strict liability. *The Journal of Legal Studies, 2*(1), 151–204.

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). Certifying and removing disparate impact (pp. 259–268). In *Presented at the proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, ACM.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS), 14*(3), 330–347.

Garcia-Martinez, A. (2017). I'm an ex-Facebook exec: don't believe what they tell you about ads. *The Guardian*. https://www.theguardian.com/technology/2017/may/02/facebook-executive-advertising-data-comment?CMP=share_btn_tw. Accessed 12 June 2017.

Garfinkel, P. (2016). A linguist who cracks the code in names to predict ethnicity. *New York Times*. http://www.nytimes.com/2016/10/16/jobs/a-linguist-who-cracks-the-code-in-names-to-predict-ethnicity.html?nytmobile=0. Accessed 6 Dec 2016.

Gettman, J., Whitfeld, E., & Allen, M. (2016). *The war on marijuana in black and white*. ACLU of Massachusetts. https://aclum.org/app/uploads/2016/10/TR-Report-10-2016-FINAL-with-cover.pdf. Accessed 11 Oct 2016.

Ghani, R. (2016). You say you want transparency and interpretability? *Machine Learning, Data Science, Analytics, Obama for America, University of Chicago, Big Data, Public Policy*. http://www.rayidghani.com/you-say-you-want-transparency-and-interpretability. Accessed 18 Oct 2016.

Gillespie, T. (2016). Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. *Information, Communication & Society, 20,* 1–18.

Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., et al. (2017). Will democracy survive big data and artificial intelligence? *Scientific American*. https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/. Accessed 12 June 2017.

Holder, E. (2014). Attorney General Eric Holder Speaks at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference. *The United States Department of Justice*. https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th. Accessed 26 October 2016.

Howard, A. (2014). Data-driven policy and commerce requires algorithmic transparency. *TechRepublic*. http://www.techrepublic.com/article/data-driven-policy-and-commerce-requires-algorithmic-transparency/. Accessed 30 July 2015.

Hu, M. (2016). *Big Data Blacklisting. Florida Law Review, 67*(5), 1735.

Introna, L. D. (2016). Algorithms, governance, and governmentality: On governing academic writing. *Science, Technology and Human Values, 41*(1), 17–49.

Johnson, D. G. (2004). Is the global information infrastructure a democratic technology? *Readings in Cyberethics, 18,* 121.

Johnson, D. G., & Noorman, M. (2014). Artefactual agency and artefactual moral agency. In P. Kroes & P. P. Verbeek (Eds.), *The Moral Status of Technical Artefacts. Philosophy of Engineering and Technology* (vol. 17). Springer, Dordrecht.

Jones, M. L. (2017). A right to a human in the loop: Legal constructions of computer automation & personhood from data banks to algorithms. *Social Studies of Science, 47*(2), 216–239.

Kharif, O. (2016). No Credit History? No Problem. Lenders are looking at your phone data. *Bloomberg.com*. https://www.bloomberg.com/news/articles/2016-11-25/no-credit-history-no-problem-lenders-now-peering-at-phone-data. Accessed 1 Dec 2016.

Kim, T. (2016). How an old hacking law hampers the fight against online discrimination. *The New Yorker*. http://www.newyorker.com/business/currency/how-an-old-hacking-law-hampers-the-fight-against-online-discrimination. Accessed 19 Oct 2016.

Kraemer, F., Van Overveld, K., & Peterson, M. (2011). Is there an ethics of algorithms? *Ethics and Information Technology, 13*(3), 251–260.

Kramer, S. (2017). An algorithm is replacing bail hearings in New Jersey. *Motherboard*. https://motherboard.vice.com/en_us/article/an-algorithm-is-replacing-bail-hearings-in-new-jersey. Accessed 11 June 2017.

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, *165*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2765268.

LaFrance, A. (2017). Uber's secret program raises questions about discrimination. *The Atlantic*. https://www.theatlantic.com/technology/archive/2017/03/uber-ghost-app/518610/?utm_source=twb. Accessed 12 June 2017.

Larson, Q. (2017). What do Uber, Volkswagen and Zenefits have in common? They all used hidden code to break the law. *freeCodeCamp*. https://medium.freecodecamp.com/dark-genius-how-programmers-at-uber-volkswagen-and-zenefits-helped-their-employers-break-the-law-b7a7939c6591#.3c6kga4q7. Accessed 12 June 2017.

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 225–258). Cambridge, MA: MIT Press.

Macaulay, T. (2017). Pioneering computer scientist calls for National Algorithm Safety Board. *Techworld*. http://www.techworld.com/data/pioneering-computer-scientist-calls-for-national-algorithms-safety-board-3659664/. Accessed 12 June 2017.

Martin, K. (2015). Ethical issues in the big data industry. *MIS Quarterly Executive, 14*(2), 67–85.

Martin, K., & Freeman, R. E. (2004). The separation of technology and ethics in business ethics. *Journal of Business Ethics, 53*(4), 353–364.

Miller, C. C. (2015). Algorithms and bias: Q. and A. With Cynthia Dwork. *The New York Times*. http://www.nytimes.com/2015/08/11/upshot/algorithms-and-bias-q-and-a-with-cynthia-dwork.html. Accessed 12 Aug 2016.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), 1–21.

Nash, K. S. (2016). Mastercard deploys artificial intelligence to pinpoint transaction fraud. *Wall Street Journal*. http://blogs.wsj.com/cio/2016/11/30/mastercard-deploys-artificial-intelligence-to-pinpoint-transaction-fraud/. Accessed 1 Dec 2016.

Newman, D. T., Fast, N., & Harmon, D. (2016). When eliminating Bias isn't fair: Decision-making algorithms and organizational justice. Presented at the Society for Business Ethics in Anaheim, CA.

Neyland, D. (2016). Bearing account-able witness to the ethical algorithmic system. *Science, Technology and Human Values, 41*(1), 50–76.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishing Group.

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, MA: Harvard University Press.

Rawls, J. (2009). *A theory of justice*. Cambridge, MA: Harvard University Press.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science, 20*(1), 1–6.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2016). Automation, algorithms, and politics| when the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. *International Journal of Communication, 10,* 19.

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society, 4*(2), 2053951717738104.

Selbst, A. D., & Barocas, S. (2018). The intuitive appeal of explainable machines. *Fordham Law Review* (**Forthcoming**).

Shaver, K. (2012). Female dummy makes her mark on male-dominated crash tests. *Washington Post*. https://www.washingtonpost.com/local/trafficandcommuting/female-dummy-makes-her-mark-on-male-dominated-crash-tests/2012/03/07/gIQANBLjaS_story.html. Accessed 20 Dec 2016.

Skeem, J. L., & Lowenkamp, C. T. (2015). Risk, race, & recidivism: predictive bias and disparate impact. *Available at SSRN*.

Smith, M. (2016). In Wisconsin, a backlash against using data to foretell defendants' futures. *The New York Times*. http://www.nytimes.com/2016/06/23/us/backlash-in-wisconsin-against-using-data-to-foretell-defendants-futures.html. Accessed 12 Aug 2016.

Sollars, G. G. (2003). A critique of social products liability. *Business Ethics Quarterly, 13*(03), 381–390.

Thornton, J. (2016). Cost, accuracy, and subjective fairness in legal information technology: A response to technological due process critics. *New York University Law Review, 91,* 1821–1949.

Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Journal on Telecommunications and High Technology Law, 13,* 203.

Urbina, I. (2013). Marijuana arrests four times as likely for blacks. *The New York Times*. http://www.nytimes.com/2013/06/04/us/marijuana-arrests-four-times-as-likely-for-blacks.html. Accessed 3 Aug 2016.

Waddell, K. (2016). How algorithms can bring down minorities' credit scores. *The Atlantic*. http://www.theatlantic.com/technology/archive/2016/12/how-algorithms-can-bring-down-minorities-credit-scores/509333/. Accessed 6 Dec 2016.

Wexler, R. (2017). How companies hide software flaws that impact who goes to prison and who gets out. *Washington Monthly*. http://washingtonmonthly.com/magazine/junejulyaugust-2017/code-of-silence/. Accessed 12 June 2017.

Winner, L. (1980). Modern technology: Problem or opportunity? *Daedalus*, *109*(1), 121–136.

Yeung, K. (2017). Hypernudge: Big data as a mode of regulation by design. *Information, Communication & Society*, *20*(1), 118–136.

Ziewitz, M. (2016). Governing algorithms myth, mess, and methods. *Science, Technology and Human Values, 41*(1), 3–16.