Are Algorithmic Decisions Legitimate? The Effect of Process and Outcomes on Perceptions of Legitimacy of AI Decisions

Kirsten Martin[1] · Ari Waldman[2]

Abstract

Firms use algorithms to make important business decisions. To date, the algorithmic accountability literature has elided a fundamentally empirical question important to business ethics and management: Under what circumstances, if any, are algorithmic decision-making systems considered *legitimate*? The present study begins to answer this question. Using factorial vignette survey methodology, we explore the impact of decision importance, governance, outcomes, and data inputs on perceptions of the legitimacy of algorithmic decisions made by firms. We find that many of the procedural governance mechanisms in practice today, such as notices and impact statements, do not lead to algorithmic decisions being perceived as more legitimate in general, and, consistent with legitimacy theory, that algorithmic decisions with good outcomes are perceived as more legitimate than bad outcomes. Yet, robust governance, such as offering an appeal process, can create a legitimacy dividend for decisions with bad outcomes. However, when arbitrary or morally dubious factors are used to make decisions, most legitimacy dividends are erased. In other words, companies cannot overcome the legitimacy penalty of using arbitrary or morally dubious factors, such as race or the day of the week, with a good outcome or an appeal process for individuals. These findings add new perspectives to both the literature on legitimacy and policy discussions on algorithmic decision-making in firms.

Keywords AI · Legitimacy · AI ethics · Algorithms · Technology ethics

Introduction

There is a new challenge to firm legitimacy—namely, the increasing use of algorithms and computers to make decisions that directly impact people's lives (Citron & Pasquale, 2014). Businesses use algorithmic decision-making (ADM) systems to make hiring, firing, and promotion decisions (Ajunwa, 2020). Banks assess loan eligibility and credit risk algorithmically (Citron & Pasquale, 2014). Social media platforms rely on ADM to moderate and curate content (Vincent, 2020; Gillespie, 2019). Recommendation algorithms nudge us toward products, news, entertainment, and information (Zuboff, 2019).

And yet, despite the increasingly frequency with which management decisions are being made algorithmically, the public is skeptical. Roughly six in ten are concerned that ADM will be biased and unfair; two-thirds think algorithmic financial scoring and hiring decisions are unacceptable (Smith, 2018). Studies show that people are largely concerned about privacy risks, biases, fairness, and usefulness of ADM (Araujo et al., 2020). They think ADM systems are less "authentic" and, therefore, less ethical than identical human decisions (Jago, 2019). And Newman et al. (2020) have shown that individuals perceive ADM used by human resources departments as reductionist and unfair. This suggests that ADM may materially impact corporate legitimacy because a firm's legitimacy is a product of and positively correlated with its actions (Chung et al., 2014; Santana, 2012).[1] Within business ethics, the perceived legitimacy of stakeholders—consumers, users, employees, etc.—are important to firms as key audiences to their

✉ Kirsten Martin
  Kmarti33@nd.edu

1 University of Notre Dame, Notre Dame, USA

2 Northeastern University, Boston, USA

---

[1] We recognize, following Eesley and Lenox (2006), that decision legitimacy merits separate evaluation and study. And, the link between decision legitimacy and overall firm legitimacy is of critical importance to business and business ethics scholars.

actions, and legitimacy is seen as a fundamental concept in business ethics (Chen et al., 2020). In their reassessment of legitimacy as a stakeholder attribute, Ali (2017, p. 157) has called identifying stakeholder perceptions of legitimacy one of three "pivotal" questions for legitimacy. In other words, understanding what is or is not legitimate is a "process that involves perceptions" (p. 160) because stakeholder and firm legitimacy is "based on the perceptions of the society, stakeholders, and management" (p. 164).

Given the risks that ADM systems pose to corporate legitimacy, this study seeks to determine whether and under what circumstances, if any, the use of ADM to make commercial decisions impacts perceived legitimacy of the decision. To be clear, this study does not seek to normatively defend the practice of using ADM to making social, economic, or policy decisions about people. Nor does it suggest that adoption of any particular form of governance makes the use of ADM legitimate in a democratic society. Rather, we build on the empirical work of Elsbach (1994), Finch et al. (2012), and Jahn et al. (2020) in that we seek to empirically measure perceptions of legitimacy of commercial actors. We do so in a novel context of ADM and also add to the literature by considering the relative importance of factors in determining legitimacy. We use the terms legitimacy *dividend* and legitimacy *penalty* to describe the positive and negative effects that the presence or absence of a condition has on perception of the legitimacy of an algorithmic decision.

Using factorial vignette survey methodology to survey individuals' normative judgments about algorithmic decisions, we ran nine surveys to measure the relative importance of governance, outcomes, and inputs on perceived ADM decision legitimacy. We find that although decision importance is negatively associated with perceptions of legitimacy, and outcomes are positively associated with perceptions of legitimacy, only particularly robust procedural governance increases perceptions of legitimacy when algorithmic decisions lead to negative outcomes or outcomes with which individuals disagree. That said, neither positive outcomes nor procedural governance can correct for the significant legitimacy penalties associated with ADM systems that use arbitrary and morally dubious factors such as race. Given that predictive ADM systems reify biases in society, our results have significant impact on firm decisions and public policy.

We contribute to legitimacy theory by explicitly testing the perceived legitimacy of ADM. Extending this type of analysis to algorithmic and computer-driven decisions, considering the degree to which the public is aware and skeptical of such decisions (Araujo et al., 2020), we extend the scope of legitimacy scholarship. First, we build on work by Elsbach (1994), Finch et al. (2012), and Jahn et al. (2020), which has explored positive drivers of decision legitimacy by demonstrating factors' relative importance.

Second, we contribute to the business ethics scholarship on legitimacy by considering it through the lens of independent variables—decision importance, procedural governance, and outcomes—that have not been studied in the business ethics literature (Badas, 2019; Gibson & Caldeiera, 2009; Tyler, 2006/1990). Finally, we contribute to public policy as regulators and firms struggle to understand how to improve ADM systems for all stakeholders. This research suggests that testing for arbitrary or morally dubious factors is more important for perceptions of legitimacy of the decision and the company than offering a better notice or impact statements, two particularly common recommendations in the literature (Kaminski, 2019a, 2019b, 2019c; Katyal, 2019; Reisman et al., 2018). Current law and new regulatory proposals prioritize procedural rules for ADM that, our study shows, will not increase perceived legitimacy.

Theoretical Context

Algorithmic Decision-Making in Firms

This study focuses on algorithmic-decision-making (ADM) systems, which, following Calo (2017), we define generally as processes involving algorithms, or sequences of logical, mathematical operations, to implement policies by software. Some ADM tools are powered by artificial intelligence of varying maturity and types; our study focuses on algorithms developed from training data generally. We also recognize that there exists a range of automation, with some decisions almost fully computerized while others are merely augmented with technology (Araujo et al., 2020; Martin, 2018). Defining all ADM systems is not our goal. As this study focuses on perceptions of legitimacy of algorithmic decisions of private, commercial firms, we assume a definition that recognizes the role of data inputs, computers, and automation of decisions.[2]

Unfortunately, researchers have shown that such algorithmic systems are risky. Because ADM systems make probabilistic predictions about the future, these programs may make mistakes about the ambiguous situations they are trying to predict; probabilities are necessarily generalized, with individual cases falling through the cracks (Eubanks, 2018; Hu, 2016). Studies have also shown that ADM's

---

[2] In conceptualizing ADM in this way, we follow a long line of scholars studying popular perceptions of artificial intelligence (AI) generally (Araujo et al., 2020; Dodge & Kitchin, 2007). This paper investigates the development of an algorithm from training data and the use of that algorithm to make a decision about individuals. We do not study the underlying technique used to develop the algorithm (neural networks, supervised or unsupervised learning, etc.). We leave that for future study.

predictive capabilities may be exaggerated (Dressel & Farid, 2018; Jung et al., 2017; Salganik et al., 2020). ADM's opacity makes algorithmic systems difficult to interrogate and hold accountable (Cheng, 2013; Loi et al., 2020). Algorithmic systems also provide an incentive for surveillance and data collection because they need large information sets for model training and analysis (Martin, 2019). This creates the circumstances for invasions of privacy and erosion of privacy norms (Zwitter, 2014).

Finally, ADM systems are biased (Johnson, 2020, n.d.; Winner, 1980). They rely on the corpus of data on which they are based and, therefore, data that is biased along race, gender, sex, and socioeconomic lines will lead to discriminatory results (Caliskan et al., 2017; Katyal, 2019; Noble, 2018; O'Neil, 2016). Developers make value-laden assumptions about how to treat missing data and outliers that also impacts the bias of the developed algorithm (Martin, 2022).

Recently, scholars have begun investigating not only the harms and discrimination created by ADM but also how individuals perceive decisions that rely on algorithms. This work has paralleled existing theory on the perceptions of fairness, trust, and legitimacy of firms and decisions.

Legitimacy

At the organizational level, entities are perceived to be legitimate when their actions are "are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Suchman, 1995, p. 574). Legitimate firms pursue "socially acceptable goals in a socially acceptable manner" (Ashforth & Gibbs, 1990, p. 177). Legitimacy is an "attitude" (Jahn et al., 2020, p. 546) or "perception" (Suddaby et al., 2017, p. 451) that is closely related to institutional credibility.

Firms seek legitimacy[3] for ethical and strategic reasons in a manner that is similar to those seeking to be trustworthy and fair. For most strategy and business ethicists, being fair, trustworthy, and legitimate is both strategically sound as well as a moral imperative and an end in itself (Freeman et al., 2020; Phillips, 1997; Pirson et al., 2017). The normative argument for firms caring about legitimacy parallels the argument for why firms should care about being perceived as trustworthy and fair. Firms should seek to understand the concerns of their stakeholders—employees, users, customers, communities—based on obligations of care (Wicks et al., 1994), fairness (Phillips, 1997), and an obligation to

create value for stakeholders (Freeman et al., 2020). As an actor within a larger industry or marketplace, firms have an obligation to be perceived as legitimate and act in a manner that is "desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Suchman, 1995, p. 574) because a firm's legitimacy impacts the perceived legitimacy of the larger industry or marketplace. In other words, a firm acting in a manner that is *not* "desirable, proper, or appropriate" within the system of values has harmful spillover effects on similarly situated firms and markets, which then harms others.

In addition, firms should care about whether their actions are perceived as legitimate for the survival of their own firm since stakeholders, including employees, suppliers, and customers, among others, rely on the firm to survive and thrive.[4] A firm that is perceived as more credible and, thus, perceived as a more legitimate company has greater customer loyalty, political leeway, is perceived to make decisions for appropriate reasons, and enjoys voluntary acceptance and compliance (Chung et al., 2014; Deephouse et al., 2017, p. 32; Finch et al., 2012; Suchman, 1995).[5]

When a firm is perceived to be legitimate, it can generate public support and loyalty (Boyd, 2000; Coombs, 1992). In addition, organizations with high perceptions of legitimacy have more flexibility in decision-making and less public and political interference. In other words, if an organization wants to maintain customer loyalty, stability within a regulatory environment, and achieve market success, it needs to be perceived as legitimate, both with respect to the organization as a whole and its individual decisions (Suchman, 1995). Those firms that make illegitimate decisions are considered illegitimate themselves, and those firms cannot survive in the market (Ruef & Scott, 1998). Therefore, businesses have powerful ethical, economic, political, and social incentives to ensure that both their individual decisions and entire organizations are perceived by the public to be legitimate.

A number of factors are known to influence perceptions of organizational legitimacy. Organizational legitimacy has

---

[3] The phrase "legitimacy of the firm" references the perceived legitimacy by others as we build on legitimacy as an "attitude" (Jahn et al., 2020, p. 546) or "perception" (Suddaby et al., 2017, p. 451). This is parallel to referring to the trustworthiness of a firm when trustworthiness is defined as the belief of others as to the trust factors of the firm (Pirson et al., 2017).

[4] While these two arguments—an overtly normative or ethical argument in one paragraph is separated from the more 'strategic' argument in a following paragraph, the authors do not see these as being so separate (Freeman, 1994; Freeman et al., 2007). However, we would like to thank a reviewer for reminding us that some readers need to see the argument made separately. We see acting ethically as good for the firm and that actions that create value for stakeholders also create value for the firm.

[5] Tyler (2006/1990) and Tyler and Huo (2002, p. 102) define legitimacy more narrowly, as "perceived obligation to comply with the directives of an authority, irrespective of the personal gains" or "a quality possessed by an authority, a law, or an institution that leads others to feel obligated to obey its decisions and directives voluntarily." Both Suchman and Tyler recognize that there is a moral valence to legitimacy.

been positively associated with conformity to norms (Suddaby et al., 2017), negatively correlated with extrinsic or self-interested motives (Jahn et al., 2020), and dependent upon transparency and communication (Elsbach, 1994). Corporate social responsibility actions and rhetoric impact perceptions of legitimacy of firms (Bronn & Vidaver-Cohen, 2009; Castello & Lozano, 2011; Du & Vieira, 2012), as do transparent executive responses to crises (Beelitz & Merkl-Davies, 2012).

Although there are different types of legitimacy in the organizational literature (Palazzo & Scherer, 2006; Suddaby et al., 2017), this study is based on the theory that individual firm decisions impact organizational legitimacy—namely, the perception that a firm is operating according to shared norms, values, and rules (Chung et al., 2014). That perception is based, at least in part, on what Boyd (2000) and Coombs (1992) call issue legitimacy—namely, the perception that a particular corporate activity is appropriate, understandable, and done for the right reasons.

Within this line of scholarship, organizational legitimacy is driven by the perceived legitimacy of an organization's specific claims, decisions, and behavior (Dowling & Pfeffer, 1975; Eesley & Lenox, 2006; Santana, 2012). To the extent that a firm uses ADM to make salient decisions about its stakeholders, the legitimacy of those decisions will have a significant impact on the legitimacy of the organization as a whole. This parallels findings that the perceived fairness of an algorithmic decisions impacts someone organizational commitment to the firm (Newman et al., 2020).

Decision Legitimacy

At the level of individual decisions, motives are also important drivers of perceptions of legitimacy in the legal studies literature on legitimacy, which mostly focuses on popular perceptions of legitimacy of institutions like the Supreme Court or the police. For instance, Badas (2019) showed that when a court hands down a decision with which one disagrees, individuals think the decision was motivated by extrinsic factors like politics and ideology and is, therefore, illegitimate. Badas (2019), Bartels and Johnston (2013), and Christenson and Glick (2015) also suggest that policy disagreements and ideology influence perceptions of the legitimacy of a given court decision and the institution of the judiciary as a whole; when a court makes a decision with which individuals disagree or negatively affects those individuals, they tend to see the court as less legitimate. At the same time, Gibson and Caldeira (2009) and Gibson et al. (2005) have shown that repeated exposure to symbols of authority create a positivity bias that insulates institutions like the Supreme Court from legitimacy penalties.

And in canonical studies of legal legitimacy, Tyler (2006/1990) and Sunshine and Tyler (2003) showed that

popular perceptions of legitimacy hinges, at least in part, on the existence of procedural safeguards and the opportunity to be heard. In Tyler's work, the legitimacy dividend of fair processes overcome any lingering distrust, opposition, or negative reaction associated with an adverse result (Tyler, 1994). That is, even those individuals who came out worse off due to the actions of authorities, institutions, or law were willing to comply with the law if the process was fair (Easton, 1965).

Similar findings have been made in organizational studies. Elsbach (1994) used a within-subject experimental design to test the cattle's industry's legitimizing strategies, finding that the public expects companies to communicate about how they are conforming to accepted norms and values. Finch et al. (2012) examined perceptions of the legitimacy of the environmentally unfriendly oil sands industry in California, finding that participants who favor environmental values judged the industry as less legitimate than those who supported economic development and outcomes. Like similar studies in the legal literature described above, these studies suggest that process—in this case, notice—and outcomes may be associated with popular perceptions with legitimacy.

Both perceptions of fairness and legitimacy have been applied to ADM. Scholars have investigated respondent-level attributes and decision factors that contribute to perceptions of fairness with ADM systems (Araujo et al., 2020; Nagtegaal, 2021; Newman et al., 2020). However, according to Kaina (2008), legitimacy is a distinct concept from fairness and trust. Within work on judgements about algorithmic decisions, nascent work on legitimacy of algorithmic decisions has suggested that governance is important to the legitimacy of ADM. For example, Danaher et al. (2017) examine the legitimacy of ADM governance mechanisms (rather than the legitimacy of the decision), and Lünich and Kieslich (2021) studied attributes of respondents (i.e. general trust) in their perception of legitimacy of a vaccine distribution algorithm. Starke and Lünich (2020) empirically examine input, process, and output legitimacy of EU decisions. What Starke and Lünich (2020) did not do, and what we attempt to do here, is empirically assess the relative importance of inputs, process, and outcomes to the perceived legitimacy of algorithmic decisions.

Hypotheses

We next develop hypotheses as to the drivers of perceived legitimacy of algorithmic decisions. Based on legitimacy and algorithmic decision scholarship, we expect the type

of decision, inputs, process, and outcomes to impact the perceived legitimacy of algorithmic decisions.

Decision Type

Social scientists studying algorithms argue that the type of decision being made algorithmically may drive norms around appropriate governance. Where Nagtegaal (2021) finds that the complexity of the decision impacts the perceptions of fair ADM, others have theorized that the degree the decision is important in the life of the individual will drive perceptions of algorithmic decisions. For example, Tufekci (2015) highlights the importance of governing "gatekeeping" algorithms; Calo (2017), Burrell (2016), O'Neil (2016), and Martin (2019) argue that "consequential" or "pivotal" decisions should be given the most attention. In other words, scholars have theorized that individuals' perception of the legitimacy of commercial use of ADM varies with their assessment of the decision's importance in their lives. This leads to our first hypothesis:

H1 As decision importance increases, individuals' perception of the legitimacy of using an algorithm to make the decision decreases.

In other words, we predict a larger legitimacy penalty for algorithmic hiring and firing decisions or access to health care decisions than for decisions about music playlists or social media curation.

Outcomes

Sociolegal scholarship on legitimacy also suggests that decision outcomes matter for popular perceptions of the legitimacy of the institutions and processes led to those outcomes. For example, Badas (2019) theorizes that disagreement with the outcome of institutional decisions negatively impact perceptions of the legitimacy of those institutions, and Gibson and Caldeira's (2009) "positivity theory" posits that agreement with the outcome an institutional decision contributes to a larger legitimacy dividend than the legitimacy penalty associated with disagreement with outcomes.

H2 With all other factors held constant, a good, positive outcome is associated with a legitimacy dividend, or an increase in the perceived legitimacy of the decision.

Arbitrary Rationales

As the algorithmic accountability literature makes clear (Noble, 2018; O'Neil, 2016), predictive algorithms based on discriminatory data and discriminatory modeling produce unjust results. For justice scholars, the idea of using arbitrary
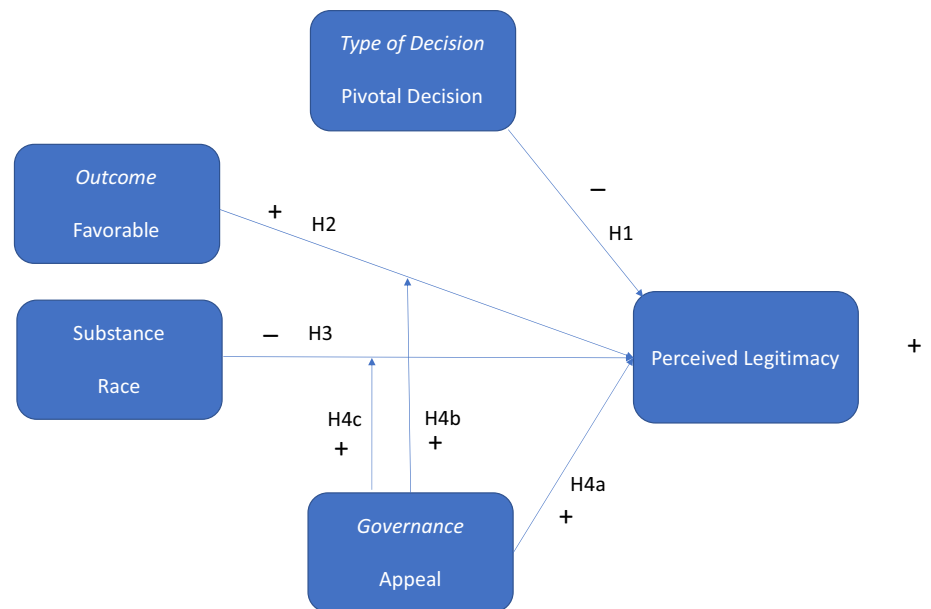
reasons or basing decisions on factors outside the context of the decision would render the decision unjust (Nozick, 1974; Walzer, 2008). For algorithmic decisions, using race-based factors in ADM contributes to illegal, and seemingly not legitimate, decisions (Barocas & Selbst, 2016). And theoretically, using race as a factor for decisions that have nothing to do with race can delegitimize decisions in law (Ellis & Diamond, 2003). The extent to which the use of arbitrary factors delegitimizes ADM has been theorized but has not been empirically tested. This leads to our third hypothesis:

H3 The use of arbitrary or race-based factors has a negative impact on perceived legitimacy.

Governance

One of the most dominant theories in legitimacy studies suggests that people perceive even adverse decisions as legitimate as long as fair procedural governance mechanisms are in place (Tyler, 2006/1990). Danaher et al. (2017) also focuses on the legitimacy of the governance mechanism of algorithmic decisions. This is the approach taken by many current and proposed data protection laws, which require organizations to communicate to customers the "logic behind" its algorithms and/or complete impact assessments (Kaminski, 2019a, 2019b, 2019c). Indeed, current policy debates center not on whether ADM requires some form of accountability mechanism, but on what that governance regime should look like (Pasquale, 2019). The theoretical argument within legal legitimacy scholarship is that robust governance—impact assessments (Kaminski, 2019b), audit trails, detailed explanations, publicly accessible code, systems testing (Citron, 2007, pp. 1305–1313), human-in-the-loop (Froomkin et al., 2019; Jones, 2017; Rahwan, 2018), impact assessments (Katyal, 2019; Reisman et al., 2018), contestability (Mulligan et al., 2020), and codes of conduct, impact assessments, and whistleblower protections (Katyal, 2019) —will legitimize decisions regardless of the outcome. Within ethics scholarship, the theorizing around governance and legitimacy is mixed. For example, de Fine Licht and de Fine Licht (2020) theorize that providing justifications for algorithmic decisions could provide grounds for perceived legitimacy, whereas, Leicht-Deobald et al. (2019) suggest that algorithm-based decisions may be perceived as *more* legitimate because it is difficult for individuals to question a complex, computerized system (Fig. 1).

In other words, the literature suggests that perceptions of legitimacy should increase as the robustness of procedural governance mechanisms increases. However, endogeneity theory (Edelman, 2016) suggests that many legal rules are developed through an endogenous process involving regulated entities themselves. If individuals perceive those rules unfair, unjust, or the product of corporate self-interest (Jahn

Fig. 1 Figure of hypotheses

Table 1 Types of decisions

Decisions	As operationalized in the vignettes
Ads	Which ads a person sees online ….their predicted preference and interests
Music	What songs are suggested in a person's playlist… their predicted preferences and interests
Video	Whether someone's posted videos are taken down online & their past and predicted behavior
Hired	Who is hired at a company….the applicant's predicted likelihood to be successful at that job
Insurance	Which insurance claims are paid …. the likelihood the policy holder will fight the claim

et al., 2020), procedural governance may not be able to provide the legitimacy dividends that corporations hope. This leads to two additional hypotheses:

H4 Any governance regime brings a legitimacy dividend to algorithmic decision-making relative to no governance, but there is a larger legitimacy dividend as procedural governance becomes more robust.

H4b Robust procedural governance mechanisms moderate the legitimacy penalties associated with bad outcomes.

Scholars have also assumed that good governance through procedural guardrails could legitimize decisions made on arbitrary or morally dubious ground (Froomkin et al., 2019; Jones, 2017; Kaminski, 2019a, 2019b, 2019c; Rahwan, 2018). However, ongoing concerns about algorithms' racial biases may call that into question. Using racist or arbitrary factors may be so delegitimizing that no amount of procedure could fix it. This leads to our final hypothesis:

H4c There is a legitimacy penalty associated with algorithms that use arbitrary or race-based factors, regardless of the form of procedural governance used.

Study Design

We used the factorial vignette survey methodology to explore the relative importance of governance, procedure, and outcomes on the perceived legitimacy of decisions based on a firm's ADM system. A factorial vignette survey presents respondents with a series of scenarios where several factors are systematically varied in the vignette; respondents then judge the scenario using a single rating task (Jasso, 2006; Wallander, 2009). The general narrative remains consistent across all vignettes with theoretically important factors (based on the hypotheses) randomly generated with replacement. The vignette factors (explained in Tables 1, 2, 3 and 4) are derived from scholarship and systematically varied in the vignette, thereby offering

Table 2 Outcome factor for vignettes

Decision	As Operationalized in the Vignettes: Bad v good outcome
Ads	a user sees ads for well-paying jobs a user sees primarily ads for poorly paying jobs
Music	a user hears songs that are totally different from those that they like. a user hears songs that are similar to those that they like
Hired	an applicant is not hired at the grocery store an applicant is hired at the grocery store
Insurance	a policy holder is then told that the claim is fully paid. a policy holder is then told that the claim will not be paid
Video	a user who posted a video then sees that the video is removed. a user who posted a video then sees that the video is not removed

Table 3 Additional factors included in algorithm

Substance	As operationalized in the vignettes
Activity	their online activity, searches, and purchases
Day	the day of the week
Race	their race and ethnicity

which the decision described was legitimate. Each respondent was presented with 30 short vignettes describing an algorithmic decision made by a firm. In general, the vignettes' narrative had four elements: a decision-maker, a decision type, the outcome of the decision, and procedural governance associated with the decision-making process, if any. The elements are illustrated in Table 1 and described in more

Table 4 Governance factor for vignettes

Governance	As operationalized in the vignettes
Human only	entire vignette is about human decision
appeals	that automated decisions can be appealed to a review board
Human governance	the organization hired a professional to oversee the automated decision
Audited	an independent audit by an external firm is conducted annually to ensure automated decisions are not biased
Impact assessment	the organization did an impact assessment on the automated system
Notification	that the organization notifies individuals that a computer program made the decision

a mechanism to test the theoretical relationships in the hypotheses (Dickel & Graeff, 2018).

This methodology allows researchers to measure the relative importance of the factors of the vignettes affect participants' attitudes, judgments, or views. The results are theoretically generalizable; the vignettes combine the benefits of laboratory experiments with those associated with experiments in the field (Oll et al., 2018).[6] As noted by Dickel and Graeff (2018), factorial vignette surveys provide a more realistic setting and triggers more honest answers than typical survey instruments (see also Auspurg et al., 2014; Weinberg et al., 2014).

In this study, decision factors were independently varied with replacement and the respondents judged the degree to

detail below. A general outline of the vignettes and samples of how they were presented to survey respondents is provided in Online Appendix A.

The use of the single rating task in the factorial vignette survey methodology supports the inductive measurement of the concept—here, the legitimacy of a company's decision— through the analysis of the factors in the vignette. Through the analysis, we measured the relative importance of the vignette factors in driving the perception of legitimacy of the respondent. This methodology, with a single rating task, has been used to measure the relative importance of vignette factors on just wages (Jasso, 2007), just punishments (Hagan et al., 2008), the trustworthiness of an organization (Pirson et al., 2017), and privacy expectations (Martin & Nissenbaum, 2020). The vignette factors constitute the theoretically important constructs that may drive the perception of rating task (trust, fairness, privacy, legitimacy). The methodology is useful to allow the researcher to identify how each factor drives the perception of the dependent variable.

---

[6] This design avoids two types of weaknesses in typical surveys. First, the respondents are forced to use a single rating task while taking into consideration multiple factors at the same time. Second, the design avoids respondent bias where respondents attempt to answer survey questions to appear more ethical or to please the researcher.

Vignette Template

The vignette template supported including different factors (below) in the vignette. The template was dynamically created as the respondent took the survey. For each survey run, a different combination of factors was included. The factors—decision type, outcome, governance, and arbitrary—and the levels of each factor are described below in Tables 1, 2, 3 and 4. The vignette template is as follows—the labels are added only for clarity as well as the number of levels or options within each factor:

> *Decision*: A computer program determines {*Decision—5 levels*}. To make the decision, the program uses {*Decision_2*}.
> *Outcome*: Based on the program, {*Outcome—2 levels*}.
> *Arbitrary*: It turns out, the decision was partly based on {*Arbitrary—3 levels*}
> *Governance*: The individual is told {*Governance—6 levels*}

Factors

Decision Type

The first set of hypotheses center on the type of decision and its importance in people's lives. We included five decision types in our study: which advertisements people see online, which songs are suggested by music app, which applicants are hired for a job, which insurance claims are filled, and which video content is taken down by an online platform. These decision types were chosen based on how pivotal they are in society based on current literature (O'Neil, 2016; Tufekci, 2015). The degree the decision was deemed pivotal was verified by a pre-test survey of 1,024 respondents on Amazon Turk. The respondents were assigned one of two conditions—either the importance to an individual or the importance to society—and asked: "Please rate the degree to which the following decision represents a critical decision affecting someone's life" or "Please rate to which the following decision represents a critical decision in society." The results are in Table A1 in Online Appendix.

Ads and music were deemed the least pivotal and hiring and insurance decisions were the most pivotal for both the individual and society based on the pre-test. In the results below, highly pivotal decisions were operationalized as vignettes with hiring or insurance decisions and low pivotal decisions were operationalized as vignettes with ads and music decisions.

Outcomes

To test the role of the outcome in the perceived legitimacy of ADM decisions, we varied the outcome from being either positive for the individual in the vignette (someone was hired) or negative for the individual (someone was not hired). The good/bad outcome was contextualized to the decision type as in Table 2.

Arbitrariness

To test hypothesis 3, we included the types of factors both arbitrary (day of the week) and discriminatory (their race and ethnicity) as well as a null (their online activities).

Governance

To test whether the type of governance over an algorithmic decision impacted the degree to which the decision is judged legitimate, we included five options for the vignettes: *transparency*, in which the organization notifies individuals that a computer program made the decision (Diakopoulos, 2020); *impact assessment*, in which the organization completed an assessment of the impact of the algorithmic process on fairness and privacy (Yam & Skorburg, 2021); *audit governance*, in which an external entity completed an annual independent audit to ensure algorithmic decisions are not biased (Mittelstadt, 2016); *human governance*, or so-called "human in the loop" of the algorithmic process (Elish, 2019); and *appeals*, in which the decision can be appealed by the individual to an internal review board (Mulligan et al., 2020). Perceptions of the legitimacy of these decisions were compared to a null in which a human made the decision without the help of an algorithm. There are, of course, other possible governance options. We chose these options because they are the most commonly proposed governance mechanisms in the legal studies literature and because they sit on a range from more robust to lax, providing a proxy for the effect on legitimacy of different types of procedural governance (Citron, 2007; Froomkin et al., 2019; Jones, 2017; Katyal, 2019; Loi et al., 2020; Martin, 2018; Rahwan, 2018; Reisman et al., 2018).

Vignette Creation

When the factors in Tables 1–4 are inserted in the vignette template, the respondent then is able to view and rate the dynamically created vignettes as shown in the examples below. The labels (Decision, Outcome, Arbitrary Governance) are only included to explain the methodology; the respondents did not see these labels.

> *Example 1*:
> *Decision*: A computer program determines what songs are suggested in a playlist (e.g., Spotify). To make the decision, the program uses their predicted preferences and interests.

Table 5 Conditions and sample

Survey	1	2	3	4	5	6	7	8	9
Vignette factors included	Decision Outcome Arbitrary Human	Decision	Decision Outcome	Decision	Decision Outcome	Decision Arbitrary Govern	Decision Outcome Arbitrary	Decision Outcome Arbitrary Govern	Decision Outcome Govern
N (respondents)	412	384	417	517	520	469	538	470	500
Vignettes	10,300	9600	10,425	12,925	13,000	11,725	16,140	14,100	15,000
Female	42.3%	41.2%	41.5%	45.7%	45.7%	40.3%	43.9%	41.3%	45.4%
Age over 35	53.8%	53.6%	49.4%	53.9%	57.1%	55.9%	54.8%	60.6%	59.0%

Outcome: Based on the program, a user hears songs that are totally different from those that they like.

Arbitrary: It turns out, the decision was partly based on their race and ethnicity.

Governance: The individual is told an independent audit by an external firm is conducted annually to ensure automated decisions are not biased.

Example 2:

Decision: A computer program determines which ads a person sees online. To make the decision, the program uses their predicted preferences and interests.

Outcome: Based on the program, a user sees ads for well-paying jobs.

Arbitrary: It turns out, the decision was partly based on the day of the week.

Governance: The individual is told that the organization notifies individuals that a computer program made the decision.

Rating Task

For each vignette, respondents were instructed to indicate on a slider the degree to which they agreed with the statement: "This decision is legitimate." The left side of the slider indicated "Strongly Disagree" and the right side of the sider indicated "Strongly Agree." The slider was on a scale of −100 to +100 with the scale not visible to the respondents. The slider option allows the respondent more freedom to rate the vignette.

Conditions and Sample

The vignette survey was run under nine conditions described in Table 5. This allowed us to isolate the importance of each factor by comparing legitimacy averages as the factors were included. For example, we tested the impact of including the outcome by comparing the averages of Survey 2 (only decision included) versus Survey 3 (decision and outcome included). Normally, a factorial vignette survey analysis would include a block analysis to isolate which factors, out of many, were dominant in driving the judgement of the

respondents. However, we were concerned that some factors, such as the arbitrary factors, could prove to be so important as to overwhelm the respondent and cloud how they perceived the other vignettes. We, therefore, spent more time running nine surveys over four months in order to better isolate the importance of each type of factor.

The surveys were run on Amazon Mechanical Turk, a crowdsourcing marketplace where researchers publish a job ("HIT") for respondents to take a survey. Each respondent rated 20–30 vignettes (depending on the condition) taking approximately 10 min; U.S. respondents were paid $1.60-$1.80 and were screened for over 95% HIT approval rate. The survey implementation was designed to minimize a number of concerns with samples from Amazon Mechanical Turk. First, the factorial vignette survey methodology was created to avoid respondent bias in normative judgments—namely, where respondents might try to game the system to appear more ethical or socially desirable. Second, the structure of the data—in two levels with individuals at the first level and vignette ratings at the second level—supports the researcher calculating whether respondents 'clicked through' without actually judging the vignette (Coppock, 2018; Daly & Natarajan, 2015; Martin, 2019; Tucker, 2014).[7] Finally, the design of the survey is to identify theoretically generalizable results as to the relative importance of factors in driving perceptions of legitimacy of AI decisions.[8]

---

[7] A few tests allow the researcher to identify whether the respondent 'clicked through' including whether the range of responses was small (clustered around −100, 0, or +100) by analyzing either the 'range' of responses or the standard deviation. These respondents were not included in the analysis. Previous studies of respondent quality comparing Mechanical Turk to a (more expensive) nationally representative survey through Knowledge Networks showed a significantly greater proportion of the sample was discarded for clicking through for Knowledge Networks (16%) compared to Mechanical Turk (2%) (Martin & Nissenbaum, 2020).

[8] Turk has been used for theoretical generalizability quite successfully, as in the examination of the relationship between concepts or ideas (Kang et al., 2014; Martin & Nissenbaum, 2017a; Redmiles et al., 2017). In critiques of Turk samples, the Turk results are compared to phone surveys (Kang et al., 2014) as well as online nation-

Table 6 Regression results of legitimacy rating task on vignette factors for each survey run

(a)

	Survey 1		Survey 2		Survey 3		Survey 4		Survey 5	
	β	p	β	p	β	p	β	p	β	p
HighPivotal	−25.68	0.00	−47.20	0.00	−10.62	0.00	−46.59	0.00	−28.25	0.00
Null = Low Pivotal										
BadOutcome	−12.92	0.00			−43.90	0.00				
Null = GoodOutcome										
DaySubstance	−31.28	0.00					−35.31	0.00		
RaceSubstance	−40.26	0.00					−46.57	0.00		
Null = activity										
AppealGovern									4.54	0.00
AuditGovern									−3.01	0.04
HumanGovern									−6.49	0.00
ImpactAssessGovern									−6.58	0.00
Null = Notice Govern										
_cons	27.30	0.00	51.35	0.00	42.10	0.00	38.71	0.00	36.23	0.00
Average	−17.11		25.53		14.14		−14.05		18.58	
N (respondents)	412		384		417		517		520	

(b)

	Survey 6		Survey 7		Survey 8		Survey 9	
	β	p	β	p	β	p	β	p
HighPivotal	−39.34	0.00	−23.30	0.00	−19.06	0.00	−6.51	0.00
Null = Low Pivotal								
BadOutcome		0.00	−15.19	0.00	−16.68	0.00	−27.43	0.00
Null = GoodOutcome								
DaySubstance	−27.29	0.00	−32.22	0.00	−22.86	0.00		
RaceSubstance	−43.69	0.00	−42.39	0.00	−39.62	0.00		
Null = activity								
AppealGovern	1.10	0.47			1.65	0.22	8.33	0.00
AuditGovern	1.40	0.36			2.15	0.11	2.96	0.03
HumanGovern	−2.78	0.07			0.28	0.83	*−0.44*	0.74
ImpactAssessGovern	−1.83	0.23			−2.02	0.13	−4.61	0.00
Null = Notice Govern								
_cons	34.92	0.00	23.62	0.00	29.03	0.00	34.34	0.00
Average	−11.38		−20.65		−9.84		19.26	
N (respondents)	469		538		470		500	

Results

Type of Decision

Hypothesis 1 states that as *decision importance increases, individuals' perception of the legitimacy of using an algorithm to make the decision decreases*. To test hypothesis 1, we regressed the rating task, the degree the decision was judged legitimate, on the vignette factors. The results are in Tables 6a and b. The coefficients (**β**) measure the relative importance of the given vignette factor to the rating task, and p is the significance of the finding. Pivotal
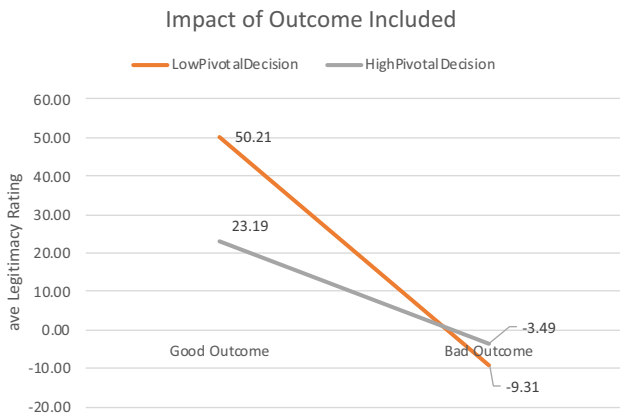

Fig. 2 Impact of good and bad outcome on pivotal decisions

decisions are those decisions deemed more important in the pre-survey test. For each survey run, more pivotal decisions lowered the legitimacy of the decision with $\beta < 0$ and $p < 0.001$. For example, in Survey 2, which only included the decision type in the vignette, including high pivotal decisions lowered the legitimacy rating by -47.20 ($p < 0.001$). This finding held across each survey and combination of factors. Less pivotal decisions made by AI are seen as more legitimate than more pivotal decisions.

Outcomes

Hypothesis 2 states *a good outcome is associated, an increase in the perceived legitimacy of the decision.* To test Hypothesis 2, we regressed the dependent variable, the degree to which the decision was judged legitimate, on the vignette factors for Survey 3 (with only the decision and outcome included). We find a bad outcome creates a legitimate penalty of -43.90 ($p < 0.001$) compared to a good outcome. In other words, a good outcome has a positive impact on perceived legitimacy. This result held even when additional factors are added in Surveys 7, 8, and 9 in Table 6b.

We also tested if the legitimacy dividend associated with a good outcome increases as decision importance increases. Figure 2 illustrates the legitimacy ratings for both good and bad outcomes. The benefit of a good outcome over a bad outcome is larger for a low pivotal decision (Good = 50.21; Bad = -9.31; ($t = 35.9793$; $p < 0.001$)). compared to a high pivotal decision (good = 23.19; bad = -3.49; ($t = 14.53$; $p < 0.001$)). In Fig. 2, the steeper slope represents a greater legitimacy penalty for bad outcomes for low pivotal decisions as compared to high pivotal decisions ($\chi^2 = 196.02$; $p < 0.001$). This could be because respondents expect a good outcome with low pivotal decisions and then penalize the surprise bad outcome more.

Arbitrary Factors

Hypothesis 3 states that *the use of arbitrary or race-based factors has a negative impact on perceived legitimacy.* The inclusion of arbitrary factors in Survey 4 lowers the legitimacy rating, all else being equal, from 25.53 (Survey 2) to -14.05 (Survey 4) ($t = -46.06$; $p < 0.001$). In general, the lowest average legitimacy rating is found in Surveys that include arbitrary factors (Survey 6 -11.38; Survey 7 -20.65; Survey 8 -9.84).

Governance and Legitimacy

Hypothesis 4a stated that the inclusion of *any governance regime has a positive impact on perceived legitimacy of algorithmic decision-making relative to no governance, but there is a larger legitimacy dividend as procedural governance becomes more robust.* To test Hypothesis 4a, we examine the regression results in Tables 6a and 6b for the surveys with governance included (Surveys 5, 6, 8, and 9). Here we see that offering an appeal process provides a legitimacy dividend for decisions with no outcome or arbitrary factors included (Survey 5). The more robust governance mechanism, to allow an appeal, positively impacts perceived legitimacy ($\beta = 4.54$, $p < 0.01$). However, alternative governance mechanisms such as an impact assessment, including a human in the loop, and having the program audited, *lower* the legitimacy of the decision compared to mere notice. This was a surprise and counter to the hypothesis.
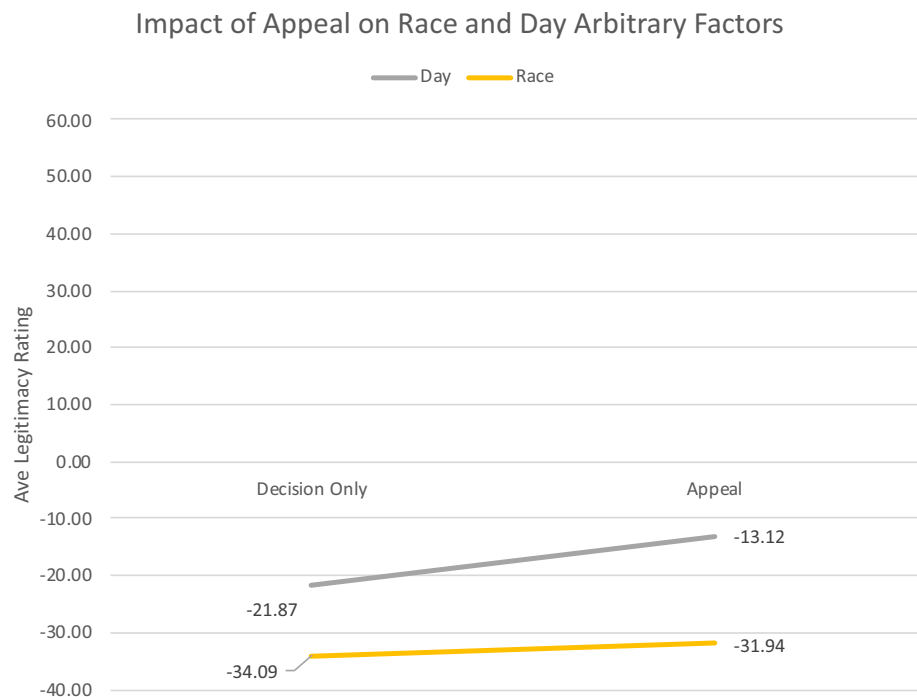
Outcomes and Governance

Hypothesis 4b states that *robust procedural governance mechanisms moderate the legitimacy penalties associated with bad outcomes.* In other words, offering an appeal to the algorithmic decision legitimizes even a bad outcome. We tested hypothesis 4b by comparing the legitimacy rating for decisions with an outcome across governance mechanisms. As shown in Fig. 3, we found that for bad outcomes, the use of an appeal process does provide a legitimacy dividend ($t = -12.53$; $p < 0.0001$); however, the use of an appeal process had no effect on the perceived legitimacy of algorithmic decisions with good outcomes ($t = -1.74$; $p = 0.04$). Further, the dividend did not 'make up' for the penalty of the bad outcome. The average legitimacy rating for a bad outcome even with an appeal process is still lower than that of a good outcome (with or without an appeal process included) —however the appeal process does close the gap between good and bad outcome legitimacy as in Fig. 3.

Fig. 3 Impact of adding a robust governance mechanism (appeal) to AI decisions with good and bad outcomes

Impact of Appeal on Good and Bad Outcomes

— Good Outcome — Bad Outcome

Good Outcome: Decision Only 33.23 → Appeal 35.83
Bad Outcome: Decision Only -4.07 → Appeal 17.89

Ave Legitimacy Rating

Fig. 4 Impact of adding a robust governance mechanism (appeal) to algorithmic decisions with race and day arbitrary factors
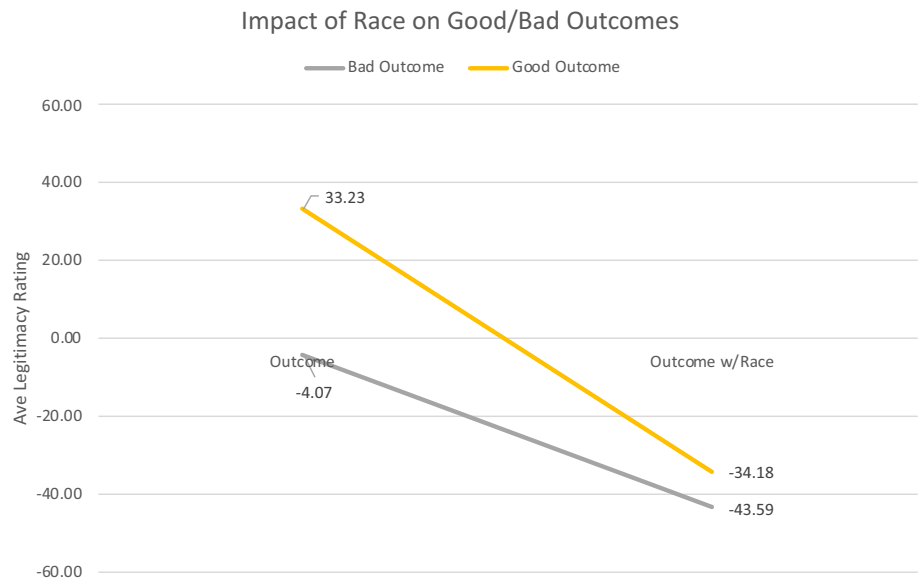
Impact of Appeal on Race and Day Arbitrary Factors

— Day — Race

Day: Decision Only -21.87 → Appeal -13.12
Race: Decision Only -34.09 → Appeal -31.94

Ave Legitimacy Rating

Arbitrary Factors and Governance

Hypothesis 4c states that *there should be a legitimacy penalty associated with algorithms that use arbitrary or race-based factors, regardless of the form of procedural governance used.* To test Hypothesis 4c, we compared the average legitimacy rating for the inclusion of race and arbitrary factors both with and without governance mechanisms. Figure 4 shows the impact of including the appeal of the decision as a governance mechanism, previously shown to have the greatest legitimacy dividend, on decisions with either race or day of the week. The impact of including an appeal on decisions that used the day of the week was modest but significant ($t = 0.3.481$; $p < 0.001$). The impact of including an appeal for decisions utilizing race was not significant ($t = 0.863$; $p = 0.19$). The legitimacy dividend of including an appeal process improves the perceived legitimacy of decisions that include arbitrary factors such as the day of the week but not when including unjust factors such as race.

Fig. 5 Impact of adding a race and day arbitrary factors to algorithmic decisions with good and bad outcomes

Impact of Race on Good/Bad Outcomes

Table 7 Summary of findings

Hypothesis	Finding
H1 As decision importance increases, individuals' perception of the legitimacy of using an algorithm to make the decision decreases	Less pivotal decisions made by AI are seen as more legitimate than more pivotal decisions
H2 A good outcome is associated with an increase in the perceived legitimacy of the decision	We find a bad outcome is a legitimate penalty compared to a good outcome. In other words, a good outcome has a positive impact on perceived legitimacy
H3 The use of arbitrary or race-based factors has a negative impact on perceived legitimacy	The inclusion of arbitrary factors in Survey 4 lowers the legitimacy rating, all else being equal, from 25.53 (Survey 2) to − 14.05 (Survey 4) ($t = -46.06$; $p < 0.001$)
H4a Any governance regime brings a legitimacy dividend to algorithmic decision-making relative to no governance, but there is a larger legitimacy dividend as procedural governance becomes more robust	The more robust governance mechanism, to allow an appeal, positively impacts perceived legitimacy. However, alternative governance mechanisms such as an impact assessment, including a human in the loop, and having the program audited, *lower* the legitimacy of the decision compared to mere notice This was a surprise and counter to the hypothesis
H4b robust procedural governance mechanisms moderate the legitimacy penalties associated with bad outcomes	For bad outcomes, the use of an appeal process does provide a legitimacy dividend. However, the use of an appeal process had no effect on the legitimacy of algorithmic decisions with good outcomes. The average legitimacy rating for a bad outcome even with an appeal process is still lower than that of a good outcome
H4c There is a legitimacy penalty associated with algorithms that use arbitrary or race-based factors, regardless of the form of procedural governance used	The legitimacy dividend of including an appeal process improves the legitimacy of decisions including arbitrary factors such as the day of the week but not when including unjust factors such as race

To better understand the legitimacy penalty for having arbitrary rationales (race or day of the week) with algorithmic decisions, we compared the impact of including the arbitrary factor of race on decisions with both good and bad outcomes. The results, in Fig. 5, show that including race as a factor is a legitimacy penalty for bad outcomes ($t = 26.40$; $p < 0.001$). In addition, any legitimacy dividends of a good outcome are erased when the decision included race as a factor ($t = 50.52$; $p < 0.001$). Adding an appeal for governance improves the legitimacy only slightly for good outcomes and race (from − 34.18 to − 24.59 with an appeal; $t = 3.14$; $p < 0.001$) and slightly for bad outcomes and race (from − 43.59 to − 33.17 with an appeal; $t = 3.32$; $p < 0.001$). Including race as an arbitrary and unjust factor in an algorithmic decision stubbornly delegitimizes the decision regardless of a good or bad outcome and regardless of including an appeal (Table 7).

Discussion and Conclusion

This paper explores the conceptual antecedents to the perceived legitimacy of algorithmic decisions. Building on cross-disciplinary empirical measurements of decision, organizational, and institutional legitimacy (Elsbach, 1994; Finch et al., 2012; Jahn et al., 2020; Tyler, 2006/1990; Badas, 2019) and theoretical accounts of legitimacy (Palazzo & Scherer, 2006; Suchman, 1995), we studied how factors of ADM—the type of decision, the outcomes, the rationale, and the governance—impact the perception of decisional legitimacy.

Empirically, we used factorial vignette methodology and made several hypotheses. We show that perceived legitimacy varies inversely with decision importance. We hypothesized that the robustness of procedural governance over algorithmic decisions made by firms would vary positively and directly with legitimacy. The data suggest something more nuanced. The only form of procedural governance that carries a legitimacy dividend is the most robust that we studied: a formal appeal to a human decision maker. Those more commonly in practice or in legislative proposals today—notice and impact assessments—either have no impact or carry a legitimacy penalty.

We hypothesized that outcomes, defined as whether an individual agrees with a decision or the decision is good for the individual, would matter more for legitimacy than procedural governance. This was mostly true, although the data suggested that there was no legitimacy dividend for positive outcomes for decisions of low importance. We also hypothesized that the use of morally questionable or arbitrary factors in making decisions would carry legitimacy penalties. Our study suggested that this was the case, erasing all legitimacy dividends associated with robust governance or positive outcomes.

Algorithms and Business Ethics

Scholars in business ethics have critically examined ADM within specific applications such as accounting (Gunz & Thorne, 2020; Munoko et al., 2020), financial services (Arthur & Owen, 2019) and HR (Leicht-Deobald, et al., 2019), and have identified specific moral implications such as social-media addiction (Bhargava & Velasquez, 2020), personalized pricing (Seele et al., 2019; Steinberg, 2020), gamification (Kim, 2018), and accountability (Martin, 2019). Previous work has also critically examined the details of algorithms in ride sharing can create gender discrimination in ratings (Greenwood et al., 2020). We extend this line of AI Ethics scholarship within business ethics to the specific examination of how ADM will impact legitimacy.

Future work could similarly examine how specific design decisions of ADM (including the use of AI and different types of machine learning) impact business ethics outcomes such as justice, fairness, trust, and legitimacy. For example, scholarship could investigate how data, governance, outcomes, and the use of AI in general impact employee perceptions of fairness. Trust in a firm, and trust factors such as ability, benevolence, and integrity, may be impacted by the use of particular data, the degree to which the firm allows decisions to be contested, and the outcome of the algorithmic decision. Scholars could leverage work on the purpose of the firm to help AI ethics scholars understand how ADM outcomes, designed by computer scientists, may undermine legitimate purposes of the firm. Rather than mistakenly seeing the introduction of algorithmic decision making as fundamentally changing how we assess ethical decisions, business ethics should leverage well researched ethical concepts to illustrate how AI does not change the nature of corporate responsibility. Work in AI Ethics within business ethics can (and should) start to connect the known moral implications of AI with our existing frameworks in business ethics in order to have concrete implications for practice as well as contribute to business ethics theory.

AI Ethics Research

Although work on fairness, accountability, transparency, and explainability (FATE) has flourished in the past few years, more work on the perceptions of ADM—parallel to the work on perceptions of human-focused decisions—should continue. For example, scholars have studied whether ADM is perceived as authentic (Jago, 2019) or fair (Araujo et al., 2020; Nagtegaal, 2021; Newman et al., 2020). More work should be done taking existing work on ethical decision making in business ethics and management to investigate whether the factors important to human decisions maintain the same relative importance to judgements about ADM. This could then help guide AI ethics research on the design of algorithmic decision systems.

This paper contributes to AI ethics scholarship by extending the moral implications that should be a concern to firms. While firms have been told to address transparency and accountability as well as discrimination, this study suggests that firms should also be concerned about legitimacy of the decision and the organization as being impacted by *how the ADM is designed*. Future work could continue to examine design decisions and their impact on different dependent variables including legitimacy, fairness, and trust in addition to how the design decisions impact users and subjects of the ADM. Extending the research to include how the firm itself is impacted by the value-laden decisions may resonate with managers and business scholarship.

Legitimacy

The results of this study add to our understanding of issue, organizational, and institutional legitimacy. Business, management, and sociolegal scholars studying legitimacy have shown that factors like communication (Elsbach, 1994), motives (Jahn et al., 2020), procedural fairness (Tyler, 2006/1990), ideology (Badas, 2019; Finch et al., 2012), and agreement or disagreement with outcomes (Gibson & Caldeira, 2009) impact legitimacy. Our study not only adds additional factors—decision importance (H1), outcomes (H2a) and the use of arbitrary or morally dubious variables (H3)—but also tests the interaction among the factors (H4a, H4b, and H4c). In so doing, we add nuance to the existing literature. For example, Tyler (2006/1990) suggests that people tend to think that authorities' decisions are legitimate when the decision-making process is fair and when individuals have the opportunity to be heard through procedural due process. Our findings are partly in line with this research, but we show that not all procedural governance mechanisms are the same. Only an appeal to a human authority is generally capable of legitimizing algorithmic systems, all other factors held constant. It is possible that other procedural mechanisms do not provide the kind of robust guarantees of fairness that individuals may associate with the right to appeal.

In work studying popular perceptions of the legitimacy of the United States Supreme Court, Badas (2019) and Gibson et al. (2005) suggest that ideological agreement or disagreement with an authority's decision is a significant driver of legitimacy, regardless of fair process. Finch et al. (2012) showed that ideological views about environmental protection and economic interests helped determine perceptions of legitimacy of a high-pollution industry. Our study tests this relationship in the context of private firm decisions and finds, with additional nuance, that outcomes do affect perceptions of legitimacy. Positive outcomes confer legitimacy dividends for decisions of high importance, but not for decisions of low importance, and negative outcomes create larger legitimacy penalties for less important decisions than highly pivotal decisions. This is in line with current literature, which suggests that individuals credit institutions that make the "right" decisions as they see them without seeking explanations (Lodge & Taber, 2013). Although our study did not examine the rationales for legitimacy penalties for outcomes mediates by decision importance, the variance may be because individuals expect that such decisions should be easy to get right and, as such, tend to react negatively when computer get them wrong.

These findings contribute to our understanding of positivity theory (Gibson & Caldeira, 2009). Positivity theory attempts to explain why individuals punish institutions less for decisions with which they disagree than they reward them for decisions with which they agree. Gibson et al.

(2014) suggest that public exposure to performative symbols of authority insulates authoritative institutions from the worst legitimacy penalties associated with ideological or policy disagreement. Our study suggests that this may only be true for decisions of high importance, but the extent, if any, of the mediation of decision importance would have to be studied empirically.

Finally, this research highlights the dangers to legitimacy for any firm that makes decisions via ADM systems that rely on race, proxies for race, or seemingly arbitrary factors that individuals do not see as related to the decision itself. Given how difficult it is to remove racial discrimination from AI systems trained on data that is itself the product of systemic and institutional discrimination, the algorithmic accountability literature is right to increasingly focus on issues of structural fairness rather than mere remediation through governance or better data (Pasquale, 2019).

Public Policy

Unfortunately, current policy proposals lag behind this the "second wave" of the algorithmic accountability literature (Pasquale, 2019). Article 22 of the European Union's General Data Protection Regulation places restrictions on private firms that employ algorithms without human intervention to make decisions that have significant effects on individuals. Firms can do so, but only if they adopt "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests," which might include "at least the right to obtain human intervention …, to express his or her point of view and to contest the decision (General Data Protection Regulation, 2018). Other provisions of the GDPR may also impact algorithmic decision-making systems, as well (Edwards & Veale, 2017), but they all share one primary common thread: they rely on procedural due process mechanisms to protect individual rights and ensure fairness (Waldman, 2021). Our study calls into question the capacity of such procedural governance to legitimize algorithmic decisions in the eyes of the public, lending credibility to arguments from critical scholars that current law needs to be more robust or, perhaps, disused in certain circumstances and for certain purposes all together (Pasquale, 2015, 2018, 2021). Although our study did not explore perceptions of the legitimacy of firms that stop using discriminatory algorithms, the strong legitimacy penalties associated with use implies the possibility of legitimacy dividends for changing course.

Of course, even the procedural guardrails of the GDPR do not exist in the United States. Firms are stepping into uncharted waters in the US, which lacks a federal agency for algorithms and neither requires notice nor impact assessments for algorithmic decisions (Crawford & Schultz, 2013). Our study suggests that US policymakers should be careful

following the GDPR's procedural governance model when, if ever, they start regulating private firm use of algorithmic decision-making systems. If, as we show, that only the most robust procedures, such as offering an appeal akin to Mulligan et al., (2020), confer legitimacy benefits, that outcomes are far more powerful drivers of perceptions of legitimacy, and that almost all kinds of algorithmic decisions are viewed as illegitimate when they use race-based or arbitrary factors, then policymakers should consider more substantive limits on algorithmic inputs and uses rather than procedural safeguards.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10551-021-05032-7.

Author Contributions All authors contributed to this manuscript.

Declarations

Conflict of interest The authors received funding from ND-Technology Ethics Center for this paper and have no known conflicts of interest. The authors have no relevant financial or non-financial interests to disclose.

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

Research Involving Human and Animal Rights This paper does include research involving humans and/or animals. The authors received IRB approval at the University of Notre Dame (#20-09-6227).

References

Ajunwa, I. (2020). The paradox of automation as anti-bias intervention. *Cardozo Law Review, 41*, 1671–1742.

Ali, M. (2017). Stakeholder salience for stakeholder firms: An attempt to reframe an important heuristic device. *Journal of Business Ethics, 144*(1), 153–168.

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society, 35*, 611–623.

Arthur, K. N. A., & Owen, R. (2019). A Micro-ethnographic study of big data-based innovation in the financial services sector: Governance, ethics and organisational practices. *Journal of Business Ethics, 160*(2), 363–375.

Ashforth, B., & Gibbs, B. (1990). The double-edge of organizational legitimization. *Organizational Science, 1*, 177–194.

Badas, A. (2019). Policy disagreement and judicial legitimacy: Evidence from the 1937 Court-Packing plan. *Journal of Legal Studies, 48*, 377–408.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*, 671–732.

Bartels, B., & Johnston, C. (2013). On the ideological foundations of Supreme Court legitimacy in the American public. *American Journal of Political Science, 57*, 184–199.

Beelitz, A., & Merkl-Davies, D. (2012). Using discourse to restore organizational legitimacy: 'CEO-speak' after an incident in a German nuclear power plant. *Journal of Business Ethics, 108*(1), 101–120.

Bhargava, V. R., & Velasquez, M. (2020). Ethics of the attention economy: The problem of social media addiction. *Business Ethics Quarterly, 31*(3), 1–39.

Boyd, J. (2000). Actional legitimation: No crisis necessary. *Journal of Public Relations Research, 4*, 341–353.

Bronn, P., & Vidaver-Cohen, D. (2009). Corporate motives for social initiative: Legitimacy, sustainability, or the Bottom Line. *Journal of Business Ethics, 87*, 91–109.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12.

Caliskan, A., Bryson, J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science, 356*, 183–186.

Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. *University of California, Davis Law Review, 51*, 399–435.

Castello, I., & Lozano, J. (2011). Searching for new forms of legitimacy through corporate responsibility rhetoric. *Journal of Business Ethics, 100*(1), 11–29.

Chen, S., Zhang, J. A., Gao, H., Yang, Z., & Mather, D. (2020). Trust erosion during industry-wide crises: the central role of consumer legitimacy judgement. *Journal of Business Ethics, 175*(4), 1–22.

Cheng, E. (2013). Being pragmatic about forensic linguistics. *Journal of Law and Policy, 21*, 541–550.

Christenson, D., & Glick, D. (2015). Chief Justice Roberts' health care decision disrobed: The microfoundations of the Supreme Court's legitimacy. *American Journal of Political Science, 59*, 403–418.

Chung, J., Berger, B., & DeCoster, J. (2014). Developing measurement scales of organizational and issue legitimacy: A case of direct-to-consumer advertising in the pharmaceutical industry. *Journal of Business Ethics, 137*, 405–413.

Citron, D. (2007). Technological due process. *Washington University Law Review, 85*(6), 1249–1313.

Citron, D., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Washington Law Review, 89*, 1–33.

Coombs, W. (1992). The failure of the task force on food assistance: A case study of the role of legitimacy in issue management. *Journal of Public Relations Research, 4*(2), 101–122.

Coppock, A. (2018). Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods, 7*(3), 613–628.

Crawford, K., & Schultz, J. (2013). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review, 55*, 93–128.

Daly, T., & Natarajan, R. (2015). Swapping bricks for clicks: Crowdsourcing longitudinal data on Amazon Turk. *Journal of Business Research, 68*(12), 2603–2609.

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., et al. (2017). Algorithmic governance: Developing a research agenda through the power of collective intelligence. *Big Data & Society, 4*(2), 2053951717726554.

Deephouse, D., Bundy, J., Tost, L., & Suchman, M. (2017). Organizational legitimacy: Six key questions. In R. Greenwood, C. Oliver,

T. Lawrence, & R. Meyer (Eds.), *The SAGE handbook of organizational institutionalism* (2nd ed.). Oxford: Sage.

Diakopoulos, N. (2020). *Transparency*. Oxford University Press.

Dickel, P., & Graeff, P. (2018). Entrepreneurs' propensity for corruption: A vignette-based factorial survey. *Journal of Business Research, 89*, 77–86.

Dodge, M., & Kitchin, R. (2007). The automatic management of drivers and driving spaces. *Geoforum, 38*(2), 264–275.

Dowling, J., & Pfeffer, J. (1975). Organizational legitimacy: Social values and organizational behavior. *Pacific Sociological Review, 18*(1), 122–136.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*(1), 1–5.

Du, S., & Vieira, E. (2012). Striving for legitimacy through corporate social responsibility: Insights from oil companies. *Journal of Business Ethics, 110*(4), 413–427.

Easton, D. (1965). *A systems analysis of political life*. Wiley.

Edwards, L., & Veale, M. (2017). Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law and Technology Review, 16*, 18–84.

Eesley, C., & Lenox, M. J. (2006). Firm responses to secondary stakeholder action. *Strategic Management Journal, 27*(8), 765–781.

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human–robot interaction (pre-print). *Engaging Science, Technology, and Society*. https://doi.org/10.2139/ssrn.2757236.

Ellis, L., & Diamond, S. S. (2003). Race, diversity, and jury composition: Battering and bolstering legitimacy. *The Chicago-Kent Law Review, 78*, 1033.

Elsbach, K. (1994). Managing organizational legitimacy in the California cattle industry: The construction and effectiveness of verbal accounts. *Administrative Science Quarterly, 39*(1), 57–88.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. Martin's Press.

Finch, D., Deephouse, D., & Varella, P. (2012). Examining an individual's legitimacy judgment using the value-attitude system: The role of environmental and economic values and source credibility. *Journal of Business Ethics, 127*(2), 265–281.

de Fine Licht, K., & de Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making: Why explanations are key when trying to produce perceived legitimacy. *AI & Society, 35*, 917–926.

Freeman, R. E. (1994). The politics of stakeholder theory: Some future directions. *Business Ethics Quarterly, 4*(4), 409–421.

Freeman, R. E., Martin, K., & Parmar, B. (2007). Stakeholder capitalism. *Journal of Business Ethics, 74*(4), 303–314.

Freeman, R. E., Parmar, B. L., & Martin, K. (2020). *The power of and: Responsible business without trade-offs*. Columbia University Press.

Froomkin, A. M., Kerr, I., & Pineau, J. (2019). When AIs outperform doctors: Confronting the challenges of a tort-Induced over-reliance on machine learning. *Arizona Law Review, 61*, 33–99.

General Data Protection Regulation. (2018). Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC, May 4, 2016, 2016 O.J. (L 119).

Gibson, J., & Caldeira, G. (2009). *Citizens, courts, and confirmation: Positivity theory and the judgments of the American people*. Princeton University Press.

Gibson, J., Caldeira, G., & Spence, L. K. (2005). Why do people accept public policies they oppose? Testing legitimacy theory with a survey-based experiment. *Political Research Quarterly, 58*, 187–201.

Gibson, J., Lodge, M., & Woodson, B. (2014). Losing but accepting: Legitimacy, positivity theory, and the symbols of judicial authority. *Law & Society Review, 48*(4), 837–866.

Gunz, S., & Thorne, L. (2020). Thematic symposium: The impact of technology on ethics, professionalism and judgement in accounting. *Journal of Business Ethics, 167*, 153–155.

Hagan, J., Ferrales, G., & Jasso, G. (2008). How law rules: Torture, terror, and the normative judgments of Iraqi judges. *Law & Society Review, 42*(3), 605–644.

Hu, M. (2016). Big data blacklisting. *Florida Law Review, 67*(5), 1735.

Jago, A. (2019). Algorithms and authenticity. *Academy of Management Discoveries, 5*, 38–56.

Jahn, J., Eichhorn, M., & Bruhl, R. (2020). How do individuals judge organizational legitimacy? Effects of attributed motives and credibility on organizational legitimacy. *Business & Society, 59*(3), 545–576.

Jasso, G. (2006). Factorial survey methods for studying beliefs and judgments. *Sociological Methods & Research, 34*(3), 334–423.

Jasso, G. (2007). Studying justice: Measurement, estimation, and analysis of the actual reward and the just reward. IZA Discussion Papers.

Johnson, G. (n.d.). Are algorithms value-free? Feminist theoretical virtues in machine learning. *Journal Moral Philosophy*.

Johnson, G. (2020). Algorithmic bias: On the implicit biases of social technology. *Synthese, 198*(10), 9941–9961.

Jones, M. L. (2017). The right to a human in the loop: Political constructions of computer automation and personhood. *Social Studies of Science, 47*, 216–239.

Kaina, V. (2008). Legitimacy, trust and procedural fairness: Remarks on Marcia Grimes' study. *European Journal of Political Research, 47*(4), 510–521.

Kaminski, M. E. (2019b). Binary governance: Lessons from the GDPR's approach to algorithmic accountability. *Southern California Law Review, 92*(6), 1529–1616.

Kaminski, M. (2019a). The right to explanation, explained. *Berkeley Technology Law Journal, 34*, 189–218.

Kaminski, M. E. (2019c). The right to explanation, explained. *Berkeley Technology Law Journal, 34*, 189.

Kang, R., Brown, S., Dabbish, L., & Kiesler, S. B. (2014). Privacy Attitudes of Mechanical Turk Workers and the US Public. In *SOUPS* (pp. 37–49). Presented at the SOUPS.

Katyal, S. (2019). Private accountability in the age of artificial intelligence. *University of California, Los Angeles Law Review, 66*, 54–141.

Kim, T. W. (2018). Gamification of labor and the charge of exploitation. *Journal of Business Ethics, 152*(1), 27–39.

Leicht-Deobald, U., Busch, T., Schank, C., Weibel, A., Schafheitle, S., Wildhaber, I., & Kasper, G. (2019). The challenges of algorithm-based HR decision-making for personal integrity. *Journal of Business Ethics, 160*(2), 377–392.

Lodge, M., & Taber, C. (2013). *The rationalizing voter*. Cambridge University Press.

Loi, M., Ferrario, A., & Vigano, E. (2020). Transparency as design publicity: Explaining and justifying inscrutable algorithms. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-020-09564-w.

Lünich, M., & Kieslich, K. (2021). Using automated decision-making (ADM) to allocate Covid-19 vaccinations? Exploring the roles of trust and social group preference on the legitimacy of ADM vs. human decision-making. arXiv preprint arXiv:2107.08946.

Lynch, J. G., Jr. (1982). On the external validity of experiments in consumer research. *Journal of Consumer Research, 9*(3), 225–239.

Martin, K. (2018). Ethical implications and accountability of algorithms. *Journal of Business Ethics, 160*, 1–16.

Martin, K. (2019). Designing ethical algorithms. *MIS Quarterly Executive, 18*(2), 129–142.

Martin, K. (2022). Algorithmic bias and corporate responsibility: How companies hide behind the false veil of the technological

imperative. In K. Martin (Ed.), *Ethics of data and analytics*. Taylor & Francis.

Martin, K., & Nissenbaum, H. (2017a). Privacy interests in public records: An empirical investigation. *Harvard Journal of Law and Technology, 31*(1), 111–143.

Martin, K., & Nissenbaum, H. (2020). What is it about location? *Berkeley Technology Law Journal, 35*(1), 251.

Mittelstadt, B. (2016). Automation, algorithms, and politics| auditing for transparency in content personalization systems. *International Journal of Communication, 10*, 12.

Mulligan, D. K., Kluttz, D., & Kohli, N. (2020). Shaping our tools: Contestability as a means to promote responsible algorithmic decision making in the professions. In K. Werbach (Ed.), *After the digital Tornado.* Cambridge: Cambridge University Press.

Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics, 167*(2), 209–234.

Nagtegaal, R. (2021). The impact of using algorithms for managerial decisions on public employees' procedural justice. *Government Information Quarterly, 38*(1), 101536.

Newman, D., Fast, N., & Harmon, D. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior & Human Decision Processes, 160*, 149–167.

Noble, S. (2018). *Algorithms of oppression*. N. Y. U. Press.

Nozick, R. (1974). *Anarchy, state, and utopia* (Vol. 5038). Basic Books.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Oll, J., Hahn, R., Reimsbach, D., & Kotzian, P. (2018). Tackling complexity in business and society research: The methodological and thematic potential of factorial surveys. *Business & Society, 57*, 26–59.

Palazzo, G., & Scherer, A. (2006). Corporate legitimacy as deliberation: A communicative framework. *Journal of Business Ethics, 66*, 71–88.

Pasquale, F. (2015). *Black box society*. Harvard University Press.

Pasquale, F. (2018). When machine learning is facially invalid. *Communications of the ACM, 61*(9), 25–27.

Pasquale, F. (2019). The second wave of algorithmic accountability. *Law and Political Economy Blog*. https://lpeblog.org/2019/11/25/the-second-wave-of-algorithmic-accountability/.

Phillips, R. A. (1997). Stakeholder theory and a principle of fairness. *Business Ethics Quarterly, 7*(1), 51–66.

Pirson, M., Martin, K., & Parmar, B. (2017). Formation of stakeholder trust in business and the role of personal values. *Journal of Business Ethics, 145*(1), 1–20. https://doi.org/10.1007/s10551-015-2839-2.

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology, 21*, 5–14.

Redmiles, E. M., Kross, S., Pradhan, A., & Mazurek, M. L. (2017). *How well do my results generalize? Comparing security and privacy survey results from MTurk and web panels to the US*. University of Maryland.

Reisman, D., Schultz, J., Crawford, K., & Whittaker, M. (2018). Algorithmic impact assessments: A practical framework for public agency accountability. AI Now Institute. https://ainowinstitute.org/aiareport2018.pdf.

Ruef, M., & Scott, W. (1998). A multidimensional model of organizational legitimacy: Hospital survival in changing institutional environments. *Administrative Science Quarterly, 43*, 877–904.

Salganik, M., et al. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *PNAS, 117*(15), 8398–8403.

Santana, A. (2012). Three elements of stakeholder legitimacy. *Journal of Business Ethics, 105*(2), 257–265.

Seele, P., Dierksmeier, C., Hofstetter, R., & Schultz, M. D. (2019). Mapping the ethicality of algorithmic pricing: A review of dynamic and personalized pricing. *Journal of Business Ethics, 170*, 697–719.

Sharpe Wessling, K., Huber, J., & Netzer, O. (2017). MTurk character misrepresentation: Assessment and solutions. *Journal of Consumer Research, 44*(1), 211–230.

Smith, A. (2018). *Attitudes toward algorithmic decision-making*. Pew Research Center.

Steinberg, E. (2020). Big data and personalized pricing. *Business Ethics Quarterly, 30*(1), 97–117.

Suchman, M. (1995). Managing legitimacy: Strategic and institutional approaches. *Academy of Management Review, 20*(3), 571–610.

Suddaby, R., Bitektine, A., & Haack, P. (2017). Legitimacy. *Academic of Management Annals, 11*(1), 451–478.

Sunshine, J., & Tyler, T. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review, 37*, 513–547.

Tucker, C. (2014). The reach and persuasiveness of viral video ads. *Marketing Science, 34*(2), 281–296.

Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Journal of Telecommunications and High Technology, 13*, 203–218.

Tyler, T. (2006/1990). *Why people obey the law*. Princeton University Press.

Tyler, T. (1994). Governing amid diversity: The effect of fair decision-making procedures on the legitimacy of government. *Law & Society Review, 28*(3), 809–831.

Tyler, T., & Huo, Y. J. (2002). *Trust in the Law: Encouraging Public Cooperation with the Police and Courts*. New York: Russell Sage Foundation.

Vincent, J. (2020). Facebook is now using AI to sort content for quicker moderation. The Verge. https://www.theverge.com/2020/11/13/21562596/facebook-ai-moderation.

Waldman, A. (2021). The new privacy law. *U.C. Davis Law Review Online, 55*, 19–42.

Wallander, L. (2009). 25 years of factorial surveys in sociology: A review. *Social Science Research, 38*(3), 505–520.

Walzer, M. (2008). *Spheres of justice: A defense of pluralism and equality*. Basic Books.

Wicks, A. C., Gilbert, D. R., Jr., & Freeman, R. E. (1994). A feminist reinterpretation of the stakeholder concept. *Business Ethics Quarterly, 4*(4), 475–497.

Winner, L. (1980). Do artifacts have politics? *Daedalus, 109*(1), 121–136.

Yam, J., & Skorburg, J. A. (2021). From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology, 23*, 611–623.

Zuboff, S. (2019). *The age of surveillance capitalism*. Profile Books.

Zwitter, A. (2014). Big data ethics. *Big Data & Society*. https://doi.org/10.1177/2053951714559253.