JCP Journal of Consumer Psychology

SCP Society for Consumer Psychology

Check for updates

Consumers Object to Algorithms Making Morally Relevant Tradeoffs Because of Algorithms' Consequentialist Decision Strategies

Berkeley J. Dietvorst and Daniel M. Bartels
University of Chicago

Accepted by Lauren Block, Editor; Associate Editor, Rajesh Bagchi

Why do consumers embrace some algorithms and find others objectionable? The moral relevance of the domain in which an algorithm operates plays a role. The authors find that consumers believe that algorithms are more likely to use maximization (i.e., attempting to maximize some measured outcome) as a decision-making strategy than human decision makers (Study 1). Consumers find this consequentialist decision strategy to be objectionable in morally relevant tradeoffs and disapprove of algorithms making morally relevant tradeoffs as a result (Studies 2, 3a, & 3b). Consumers also object to human employees making morally relevant tradeoffs when they are trained to make decisions by maximizing outcomes, consistent with the notion that their objections to algorithmic decision makers stem from concerns about maximization (Study 4). The results provide insight into why consumers object to some consumer relevant algorithms while adopting others.

Keywords Morality; Judgment; Decision-making; Values; Consequentialism

More than ever before, consumers can use algorithms (any tool that uses a fixed step-by-step decision-making process, including statistical models, actuarial tables, and computer programs) to make or augment their decisions. Consumers can, for example: (a) use recommender systems to choose between products, (b) employ algorithmic dating websites to find a partner, (c) use online calculators to determine how much to save for retirement or spend on a house, (d) consult online mapping services to determine how to get from one place to another, and (e) offload investment decisions to robo-advisors. Even people who choose not to use any of these services are still affected by algorithms in commerce. Organizations use algorithms to decide (a) how to price their products, (b) whether or not to approve consumers for loans, (c) which consumers receive advertisements and promotions, and (d) which people are awarded scarce resources (e.g., scholarships and organ donations).

People have embraced many of these algorithms. Hundreds of millions of consumers use music streaming services, like Pandora and Spotify, which offer algorithmic recommendations as a major selling point (Dunn, 2017). Over one billion people use Google Maps instead of more traditional methods of navigation (Perez, 2016). Millions of Americans use online dating websites, which use algorithms to find potential partners (Thottam, n.d.).

On the other hand, consumers have been reluctant to accept algorithmic decision makers in some domains. For example, people prefer not to use algorithms for medical decisions (Longoni, Bonezzi, & Morewedge, 2019) and prefer doctors who do not consult algorithms before making a medical decision (Shaffer, Probst, Merkle, Arkes, & Medow, 2013). Consumers are generally hesitant to use algorithms for tasks that seem subjective (Castelo, Bos, & Lehmann, 2019). When it comes to customer service, people experience greater discomfort when interacting with human service robots than humans (Mende, Scott, van Doorn, Grewal, & Shanks, 2019) and are less likely to purchase items from chatbots than human workers (Luo, Tong, Fang, & Qu, 2019). Additionally, consumers sometimes prefer recommendations from human agents over recommendations from algorithmic recommender systems (Yeomans, Shah, Mullainathan, & Klienberg, 2019), and people prefer investment decisions that have

Received 15 December 2020; accepted 27 June 2021
Available online 07 July 2021

Berkeley J. Dietvorst is an associate professor of marketing at The University of Chicago Booth School of Business. Daniel M. Bartels is a professor of marketing at The University of Chicago Booth School of Business.

Correspondence concerning this article should be addressed to Berkeley J. Dietvorst, The Booth School of Business, University of Chicago, 5807 S. Woodlawn Ave., Chicago, IL 60637. Electronic mail may be sent to berkeley.dietvorst@chicagobooth.edu.

© 2021 Society for Consumer Psychology
All rights reserved. 1057-7408/2022/1532-7663/32(3)/406–424
DOI: 10.1002/jcpy.1266

moral undertones to be made by human fund managers instead of algorithms (Niszczota & Kaszás, 2020).

Complicating matters, there seems to be substantial heterogeneity in consumers' tolerance for algorithmic decision makers—not all people react the same way to a specific algorithmic decision maker. For example, Leung, Paolacci, and Puntoni (2018) find that consumers who have identity motives for their consumption decisions often resist automated features in products, while those who do not have identity motives are more accepting of automation. People's financial status affects their willingness to interact with algorithmic products that have anthropomorphized features (Kim & McGill, 2018). Many consumers report that they would be uncomfortable riding in a self-driving car (Leinert, 2018), while others are eager to adopt this technology. Further, consumers do not want all self-driving cars to operate in the same manner; people want the self-driving cars that they ride in to protect passengers at all costs, while they want the self-driving cars that others ride in to sacrifice passengers for the greater good (Bonnefon, Shariff, & Rahwan, 2016).

In this research, we contribute to the growing literature on consumers' adoption of automation (see Castelo et al., 2019; Dietvorst & Bharti, 2020; Dietvorst, Simmons, & Massey, 2015, 2016; Kim & McGill, 2018; Leung et al., 2018; Longoni et al., 2019; Luo et al., 2019; Mende et al., 2019) by investigating a factor that can help to explain why the acceptability of automation varies between different domains and consumers. Specifically, we find that consumers are less tolerant of a company using an algorithm to make a decision when they feel that the decision poses a morally relevant tradeoff. Tradeoffs are decisions for which not all desirable outcomes can be achieved simultaneously, for example, where more of one good thing means less of another. In other words, for these decisions, different considerations need to be balanced against each other. For our purposes, tradeoffs are morally relevant when they entail potential harm and/or the limitation of one or more persons' resources, freedoms, or rights (Turiel, 1983). As examples, we propose that consumers are unlikely to adopt an algorithmic product that makes tradeoffs that they feel to be morally relevant (e.g., a self-driving car choosing whether to prioritize the safety of passengers or pedestrians) and may choose not to do business with an organization that uses algorithms to make tradeoffs that they feel to be morally relevant (e.g., balancing profit against the well-being of customers).

Our work also investigates why consumers express these preferences. We propose that people expect algorithms to make decisions in a more consequentialist manner than human decision makers —they expect algorithms to focus on decision outcomes instead of decision processes and other considerations. Further, we propose that people find this consequentialist decision process to be objectionable when they feel that the tradeoff in question has moral content. Thus, we also contribute to the literature on consumer moral decision-making (see Bhattacharjee, Berman, & Reed, 2013; Campbell & Winterich, 2018; Ehrich & Irwin, 2005; Irwin & Naylor, 2009; McGraw, Schwartz, & Tetlock, 2011; Reczek, Irwin, Zane, & Ehrich, 2017) by documenting a decision strategy that consumers find to be objectionable for morally relevant tradeoffs.

Theoretical Development

We propose that people expect algorithms to make decisions in a different manner than human decision makers. Specifically, they may expect algorithms to approach decisions as maximization problems. For our purposes, "maximization" refers to a decision process where the available predictive information is used to make a decision that is expected to maximize (or minimize) a predefined outcome of interest. In these kinds of processes, one or more outcomes are selected, and the process makes the best use of the available information to maximize the goodness of the result with respect to those outcomes. For example, a maximizing strategy for the game of blackjack would maximize the player's expected earnings (the predefined outcome of interest) by choosing the action (hit, stand, split, and double) that leads to the greatest expectation of earnings given the available information (the value of all known cards).

Many popular algorithms that are commonly used in consumer domains are based on maximization. Ordinary least squares regression determines the coefficients on input variables by finding the minimum sum of squared errors—the difference between the prediction and the outcome. Neighborhood-based collaborative filtering algorithms (which companies use in recommender systems) often operate by finding the item that has the highest rating among users who are similar to the customer (Ricci, Rokach, & Shapira, 2011). Decision trees used for classification categorize targets by finding the partitions in the data that maximize the difference between the resulting groups regarding the classification of interest (Lior, 2014). The decision process that each of these algorithmic methods chooses is completely dictated by their maximization (or minimization) goal; the decision process that best accomplishes this goal is always chosen, and all other considerations are disregarded.

Consumers' experience with algorithms in the marketplace likely reinforces the notion that algorithms usually approach a decision by maximizing some concrete objective. For example, services that give directions like google maps find the route that minimizes the expected time to travel from point A to point B, dating services like eHarmony display potential partners who are the highest rated on some type of compatibility score, and Google displays the web pages that are scored as being the most relevant to a consumer's search. Thus, consumers are likely to have seen evidence in the marketplace that algorithms operate by maximizing outcomes.

> *H1*: Consumers believe that decisions made by algorithms have more of a basis in maximization than decisions made by humans.

The notion that algorithms make judgments through the lens of maximizing outcomes means that their decision processes share important properties with the philosophical concept of consequentialism in normative ethics. Consequentialism is the theory that the morality of an act can be judged solely by its consequences (Kagan, 1998). According to this theory, a moral decision maker should make choices by weighing the consequences of choosing each option and then selecting the option that leads to the best expected outcomes. Importantly, according to consequentialism, only decision outcomes matter—the morality of an act is not affected by its decision process or other considerations. For example, utilitarianism, a well-known type of consequentialism that argues that the morally right action is the one that leads to the best consequences with respect to welfare interests (Kagan, 1998), contends that moral decision makers must make choices that maximize total well-being. Because algorithms are based on maximization, they operate in a similar manner as a consequentialist—algorithms' decision processes are dictated by maximization of outcomes. Algorithms will always make the choice that is expected to generate the best outcome on some criterion (or combination of criteria), and algorithms do not consider factors that will not impact the focal criterion (or criteria) in expectation. In other words, algorithms do not consider how a decision or estimate is reached, they only try to generate the best outcome in expectation—which is exactly how consequentialists approach morally relevant tradeoffs.

The literature on moral decision-making suggests that many people may find maximization as a decision process to be objectionable in morally relevant domains. Many moral philosophers have criticized consequentialism as a normative approach to moral decision-making, have constructed scenarios demonstrating consequentialism's pitfalls (Kamm, 2000), and oppose using consequentialist theories, such as utilitarianism, as normative benchmarks for moral decision-making. Further, most ethical philosophers are not utilitarians themselves; a survey of 73 professors with PhDs in philosophy who specialize in ethics found that only 27% endorsed principles of utilitarianism (compared to 37% and 22% endorsing principles of deontology and virtue ethics, and 14% endorsing none of the above (personal communication: reanalysis of data reported in Schwitzgebel & Cushman, 2012).

Importantly, resistance to consequentialist decision-making extends to laypeople. Work on moral value tradeoffs finds that lay decision makers often prioritize upholding moral rules—for example "deontological rules" (see Davis, 1993, for a description of contemporary deontological theory), like "do no harm," which pertain to actions or processes themselves and not the downstream consequences of those actions over optimizing outcomes (see Bartels, Bauman, Cushman, Pizarro, & McGraw, 2016). Further, people judge others who make deontological judgments as more moral and trustworthy than others who make consequentialist judgments (Everett, Pizarro, & Crockett, 2016). Finally, consumers who adopt a maximizing mindset are more likely to commit immoral acts (Goldsmith, Roux, & Ma, 2018).

We believe that maximization as an approach will face two specific challenges when applied to morally relevant tradeoffs. First, maximization entails selecting one outcome (or combination of outcomes) to maximize; however, there is often no general "right" or "best" approach to morally relevant tradeoffs, and in fact, there is often lack of agreement over how to make decisions involving morally relevant tradeoffs. To complicate matters further, in practice, moral decisions are often quite complicated (see Bartels et al., 2016; Bennis, Medin, & Bartels, 2010), and people's preferences in morally relevant scenarios are often very sensitive to context (Bartels, 2008). As a result, people often have preferences that vary considerably across similar morally relevant scenarios, and many people's preferences cannot be adequately characterized by assuming they are pursuing any one objective across (sometimes, seemingly similar, to the outside observer) contexts (see Bartels et al., 2016). In contrast, maximization requires that one predetermined goal guides decision-making across all tradeoffs. That is, maximization requires approaching all tradeoffs with the same objective, while human decision makers often select an objective based on the specific details of the tradeoff at hand. We believe that people may object to maximization for morally relevant tradeoffs because of this contrast between maximization's singular approach to morally relevant tradeoffs and people's flexible approach to these same decisions.

Given this reasoning, consumers may find maximization to be an objectionable decision process even in the best-case scenario when this decision strategy has a noble goal like maximizing welfare. However, many of the algorithms employed by companies in marketing contexts are likely to have less virtuous allocentric goals, and this may make maximization even less tolerable to consumers as a result. For example, companies may have the goal of maximizing profit, impressions, engagement, or some other objective.

Second, people may object to maximization as an approach to morally relevant tradeoffs because maximization strategies may likely permit objectionable acts in pursuit of their maximization goal. People often evaluate the morality of decisions based on the actions of the decision maker (see Bartels et al., 2016). For example, work on procedural justice shows that people's satisfaction with allocation decisions often depends on the perceived fairness of the decision process in addition to the decision outcomes (Alexander, & Ruderman, 1987; Konovsky, 2000; Lind & Tyler, 1988; Rhoades & Eisenberger, 2002; Tyler, 1988). This means that there are many actions that people deem unacceptable, even if they lead to better outcomes. For example, people find it objectionable to compromise on protected values, which stem from deontological rules like "tell the truth" and "do no harm," regardless of the consequences of the decision (Baron & Spranca, 1997). Similarly, people find it objectionable to balance the costs and benefits of options when a decision maker faces a "taboo tradeoff," where sacred values, like lives or natural resources, are weighed against secular goods, like money (Deghani et al., 2009, 2010; Tetlock, 2002; Tetlock, Kristel, Elson, Green, & Lerner, 2000).

In contrast, a maximizing decision process optimizes the (prespecified and targeted) consequences of decisions without regard for the way those outcomes are produced. So, algorithms based on maximization do not naturally consider whether the actions involved in carrying out the decision violate any moral rules and are comparatively likely (i.e., more likely than other decision processes) to violate these rules as a result. Moreover, it may be infeasible to constrain the maximization process to avoid violating any moral rules because there are a multitude of potentially relevant moral rules to consider for any morally relevant tradeoff (see Nozick, 1974). Further complicating matters, these moral rules often conflict, and people do not always prioritize them in ways that are easy to describe or predict (Bartels, 2008).

Compounding these concerns, algorithms based on maximization are often not transparent. For example, even experts often do not understand exactly how a specific machine learning algorithm arrives at a decision (Burrell, 2016; Kroll et al., 2016), or cannot explain why a machine learning algorithm operates the way that it does (Kroll et al., 2016). So, it may be hard or impossible to tell whether an algorithm adopts a decision process that is likely to violate moral rules. Considering these features of maximization as a decision process, we propose that consumers may find the premise of making decisions with algorithms that use maximization to be morally dubious when they feel that the tradeoff in question is morally relevant.

> *H2a*: Consumers who feel that a particular decision domain is more morally relevant will express more intolerance for consequentialist algorithms operating in that domain than those who feel that the domain is less morally relevant.

Following the same logic laid out above, we propose that consumers will more strongly object to algorithmic decision makers in decision domains that have more moral relevance. Previous research has suggested that people may be reluctant to use algorithms to make morally relevant decisions; however, this previous work has not provided evidence that people more strongly object to algorithms making decisions in domains with more moral relevance. For example, Bigman and Gray (2018) found that within an ostensibly moral decision domain, people rate human decision makers as more permissible than algorithmic decision makers, but do not demonstrate that this tendency differs between domains. Jago (2019) found that across five different domains, participants believed that human decisions would produce more ethical outcomes than algorithmic decisions; however, this tendency did not significantly differ between the domains. From this previous research, it is not clear that people specifically object to algorithms operating in morally relevant domains because those projects did not manipulate or measure the moral relevance of the decision domain. Thus, there is currently no clear evidence that consumers more strongly object to algorithmic decision makers in more morally relevant domains.

Our work builds on this previous work by testing whether consumers express stronger objections to algorithms in more morally relevant domains. Critically, we manipulate the decision domain both between (Study 2) and within (Studies 3a & 3b) participants and test whether participants' feelings of the moral relevance of a decision mediate the effect of the decision domain on their tolerance for algorithmic decision makers. This allows us to test whether the moral relevance of a domain affects people's tolerance for algorithmic decision makers operating in that domain. Further, we demonstrate the relevance of this proposal for consumer behavior by using consumer relevant domains and dependent variables.

> *H2b*: Consumers will be less tolerant of algorithmic decision makers in decision domains that are more morally relevant.

As previously described, we propose that the specific reason why consumers object to algorithms in morally relevant domains is that they feel that maximization is an inappropriate approach for morally relevant tradeoffs. Thus, we expect that participants' feelings about the appropriateness of maximization as a decision-making strategy in a domain will relate to their tolerance for algorithmic decision makers. Specifically, we propose that consumers' intolerance for algorithms in morally relevant domains can be at least partially attributed to their objections to using maximization to make morally relevant tradeoffs.

> *H3*: Consumers are less tolerant of algorithms making tradeoffs in more morally relevant domains because they feel that maximization is an inappropriate approach for making tradeoffs in more morally relevant domains.

Following H3, our theorizing suggests that consumers are more tolerant of human decision makers in morally relevant domains because they expect human decision makers will be less likely to use maximization as a decision strategy. Given this reasoning, if people have reason to believe that a human decision maker will use maximization as a decision strategy for a morally relevant tradeoff, they should express intolerance for that human decision maker just as they would with an algorithmic decision maker. On the other hand, informing consumers that an algorithm will use maximization for a morally relevant tradeoff may have less of an impact because people naturally assume that algorithms make decisions using maximization. Thus, we expect that informing consumers that a decision maker will be trained to use maximization for a morally relevant tradeoff will have a greater negative impact on their tolerance for human than algorithmic decision makers.

> *H4*: Informing consumers that a decision maker is trained to use maximization to make a morally relevant tradeoff will moderate the effect of the decision maker being a human versus an algorithm on consumers' tolerance for that decision maker.

Material and Methods

The current studies test the hypotheses laid out above. In Study 1, participants rated a human's and an algorithm's likelihood of using maximization as a decision strategy in one of seven domains as a test of Hypothesis 1. In Study 2, participants reported their tolerance for a company starting to use an algorithmic decision maker in one of seven domains that varied in moral relevance as a test of Hypotheses 2a, 2b, and 3. In Studies 3a and 3b, participants read about six different types of tradeoffs where resources are allocated to consumers and expressed their preference for a human versus an algorithm allocator as an additional test of Hypotheses 2a and 2b. In Study 4, we manipulated whether participants read that a human or algorithmic decision maker was trained to make a morally relevant tradeoff using maximization as a test of Hypothesis 4.

We preregistered all studies before collecting any data. We report all exclusions (if any), all manipulations, and all measures in the manuscript. In the online supplemental materials [https://osf.io/4m8ba/], we post data for all studies, code for all analyses, the original materials for all studies, preregistrations for all studies, and additional supplements describing the studies and data.

Study 1

In Study 1, we investigate Hypothesis 1; consumers believe that decisions made by algorithms have more of a basis in maximization than decisions made by humans. Participants were assigned to read about one of seven different tradeoffs that a health insurance company makes and rated a humans' and an algorithms' likelihood of using maximization to make that tradeoff.

Sample & Procedure

We preregistered that we would recruit 700 participants from Amazon Mechanical Turk. After posting the study, 777 participants clicked on the study link, 46 participants failed an attention check and were not allowed to begin the study, and 707 participants completed all dependent measures. This sample was 63% male and 36 years old, on average.

Participants who passed the attention check learned that they would read about a decision that a health insurance company makes and then answer some questions about that decision. Next, participants were assigned to read about one of seven different tradeoffs. The seven tradeoffs entailed the insurance company determining: how much a customer's premium will cost (Premium), whether or not to cover a medicine that has been prescribed to one of their customers (Prescription), which doctors to recommend when a patient asks for a list of covered doctors of a certain specialty (Recommendation), which customers they will send advertisements to (Advertisement), whether or not a customer qualifies for a specific insurance plan (Plan), which insurance agent gets assigned to a customer (Agent), and when to ask customers if they would like to renew their plan for the following year (Renewal). Pilot studies and the results of Study 2 (see the second paragraph of the "Moral conviction" section of the results) find that people find these seven domains to significantly differ in the degree to which they are morally relevant.

Participants were randomly assigned to imagine that the company has chosen to use either a human employee or an algorithm to make their assigned decision. Next, participants rated the extent to which the human employee and algorithm would approach the decision using maximization in randomized order. Participants made these ratings of either the human or algorithm on a three-item scale that we developed in pretests (see Supplement 1 in the Supplemental Materials document for details). These questions asked to what extent: (a) The employee (algorithm) is likely to make this decision by only weighing measurable costs and benefits, (b) The employee (algorithm) is likely to make this decision by maximizing concrete measures of outcomes regardless of all other considerations, and (c) The employee (algorithm) is likely to make this decision by considering data alone. Participants expressed their agreement with each statement on a five-point Likert scale, with (1) being "Strongly agree" up to (5) being "Strongly disagree." Following their ratings, participants were asked to imagine that the entity they had not considered (either the human or the algorithm) would make their assigned decision and answered the same three questions. Finally, participants reported their age, sex, and highest completed level of education and completed the survey.

Results

First, we investigate the three items that measured propensity to use maximization as a decision strategy. The Cronbach's alpha of these three items was 0.742 when measuring the algorithm's propensity to use maximization and 0.751 when measuring the human's propensity to use maximization. Thus, we combine these items to create a measure of propensity to use maximization. For each of the seven decisions, participants reported that they believed the algorithm was more likely to use maximization as a decision strategy than a human decision maker, $ts(\geq 99) \leq -5.65$, $ps < .001$, (see Table 1). On average, participants' ratings of the algorithm's likelihood to use maximization were 0.83 higher (on a 5-point scale) than the ratings of the human, and this difference did not significantly differ by condition, $F(6,700) = 0.57$, $p = .76$. These results support Hypothesis 1, as participants believed that decisions made by an algorithm would have more of a basis in maximization than decisions made by a human employee across a variety of domains.

Discussion

The results of Study 1 suggest that consumers assume that algorithms employed by organizations may take a different approach to decision-making than human employees; specifically, they are more likely to use maximization. Study 2 investigates the implications of this belief for consumers' tolerance of algorithmic decision makers.

Table 1
Participants' Ratings of Propensity of the Human and Algorithm Using Maximization as a Decision Strategy

	Human			Algorithm				
Domain	*N*	*M*	*SD*	*N*	*M*	*SD*	*t*	*p*
Advertisement	101	3.14	0.89	101	3.95	0.88	6.55	<.001
Agent	100	3.12	0.88	100	4.11	0.75	7.99	<.001
Plan	101	3.23	1.11	101	3.99	0.97	6.12	<.001
Premium	101	3.15	0.98	102	4.06	0.87	7.41	<.001
Prescription	101	3.10	0.94	101	3.86	1.00	6.26	<.001
Recommendation	102	3.18	1.04	102	3.92	0.94	6.01	<.001
Renewal	101	3.17	0.98	101	4.03	0.97	7.04	<.001

Study 2

In Study 2, we test the hypothesis that people object to algorithms operating in morally relevant tradeoffs because of their objections to consequentialist decision strategies in these domains. Participants were assigned to read that an insurance company was going to start using an algorithm to make one of the seven decisions described in Study 1. Participants rated the moral relevance of the decision in question, their propensity to switch to a different insurance company if they had been a customer, and their feelings about whether or not making the decision they read about using maximization was appropriate. We investigate whether participants reported that they would be more likely to switch insurance companies if they were assigned to consider a decision with greater moral relevance (H2b), whether participants who felt that their assigned decision had greater moral relevance reported a greater likelihood of switching companies (H2a), and how participants' ratings of the acceptability of maximization relate to their switching intentions (H3).

Sample & Procedure

We preregistered that we would recruit 700 participants from Amazon Mechanical Turk. After posting the study, 858 participants clicked on the study link, 92 participants failed an attention check and were not allowed to begin the study, and 706 participants completed all dependent measures. This sample was 51% male and 36 years old, on average.

Participants who passed the attention check learned that they would read about a decision that a health insurance company makes and then answer questions about that decision. Next, participants were assigned to read about one of the seven tradeoffs used in Study 1. Next, participants rated the moral relevance of their assigned decision. We used an established scale of "moral conviction" to measure the moral relevance of the decision adapted from Skitka, Bauman, and Lytle (2009), which is commonly used to measure "perceptions of morality and immorality, right and wrong" (Skitka, 2010, p. 268). The questions asked: "To what extent are your feelings about deciding [decision domain] a reflection of your core moral values and convictions?", and "To what extent are your feelings about deciding [decision domain] deeply connected to your beliefs about 'right' and 'wrong'?" on two five-point scales (ranging from "not at all" to "very much"). We average these two ratings to create a measure of moral conviction.

On the next page, participants read that the insurance company has traditionally used a human employee to make the target decision; however, they are considering using an algorithm to make this decision in the future. Participants then completed the dependent variable; they were asked to imagine that they were a customer of the insurance company and reported their likelihood of switching to a new insurance provider (switching intentions) if the company started using an algorithm to make the target decision. Participants reported their response on a five-point Likert scale with the labels: 1—No, I would not switch, 2—There is a slight chance I would switch, 3—There is a moderate chance I would switch, 4—There is a good chance I would switch, and 5—Yes, I would switch.

On the next page, participants completed a scale measuring whether they believed it was appropriate to make their assigned decision with maximization using questions based on the scales from Study 1. Participants reported whether or not they agreed with three items that read: "It is acceptable for the insurance company to make this decision by only weighing measurable costs and benefits," "It is acceptable for the insurance company to make this decision by maximizing concrete measures of outcomes regardless of all other considerations," and "The insurance company can make acceptable decisions by considering data alone." These items were presented in randomized order. Participants expressed their agreement with each statement on a five-point Likert scale, with (1) being "Strongly agree" up to (5) being "Strongly disagree". Finally, participants reported their age, sex, and highest completed level of education and completed the survey.

Results

Moral conviction. The two questions we used to measure participants' moral conviction for their assigned condition had a Cronbach's alpha of 0.87, and we combined these measures to create a scale of moral conviction. In support of Hypothesis 2a, participants' ratings of moral conviction and switching intentions were significantly correlated, $r(705) = .42$, $p < .001$; participants who felt more moral conviction for their assigned decision reported stronger switching intentions. Further, this significant relationship persists after accounting for the decision domain that the participant was assigned to; participants' ratings of moral conviction are a significant predictor of their switching intentions in an OLS regression with six dummies controlling for participants' assigned decision domain, $b = .32$, $t(699) = 8.62$, $p < .001$. These results support the notion that within a decision domain, the more moral conviction that a consumer feels for the tradeoff at hand the less tolerant they will be of using an algorithm to make that tradeoff.

In support of Hypothesis 2b, participants were also less tolerant of algorithms in domains that that were more morally relevant on average. Participants felt that some of the decision domains were more morally relevant than others; after running a regression of participants' moral conviction ratings on decision domain dummies with no constant, a test of the hypothesis that the seven domain dummies are equal was rejected, $F(6, 701) = 22.84$, $p < .001$. Additionally, participants expressed different switching intentions across the seven conditions; after running a regression of participants' switching intentions on decision domain dummies with no constant, a test of the hypothesis that the seven domain dummies are equal was rejected, $F(6, 700) = 26.75$, $p < .001$. What's more, participants expressed stronger switching intentions in the domains that were more morally relevant; participants' aggregate ratings of each decision domain in terms of moral conviction and switching intentions are significantly correlated, $r(5) = .93$, $p = .002$ (see Table 2). These results are consistent with the notion that participants were less tolerant of algorithms in more morally relevant domains.

Further, participants' ratings of moral conviction mediate the effect of their assigned decision domain on their switching intentions. As noted, participants expressed significantly different switching intentions across the seven domains, $F(6, 700) = 26.75$, $p < .001$. After accounting for participants' moral conviction ratings by adding them to this regression, the F-value drops 12.57 points, $F(6, 699) = 14.18$, $p < .001$, and a bootstrapped confidence interval of this drop excludes zero (8.36, 17.76). This mediation is consistent with the notion that participants were less tolerant of algorithms in some domains because those domains were more morally relevant.

Acceptability of maximization. Next, we turn our attention to the scale measuring the acceptability of using maximization. Consistent with the results of Study 1, a factor analysis confirmed that all three items load on the same factor and have a Cronbach's alpha of .797. Thus, we average participants' responses to these three items to create a scale of acceptability of maximization.

Participants' ratings of the acceptability of using maximization relate to their switching intentions; participants expressed stronger switching intentions the more they felt that maximization was unacceptable within their assigned decision domain, $r(704) = .41$, $p < .001$. In the aggregate data, participants expressed stronger switching intentions in domains where they felt that it was more unacceptable to use maximization (see Figure 1), $r(5) = -.96$, $p < .001$.

Consistent with H3, we find that participants' feelings of moral conviction and rated acceptability of using maximization explain common variance in their switching intentions. This finding is consistent with the notion that participants express greater switching intentions when they feel greater moral conviction for a decision because they feel that maximization is a less acceptable decision strategy. As noted above, participants' feelings of moral conviction mediate the effect of their assigned decision domain on their switching intentions; the F-value of the null hypothesis test that participants' switching intentions are equal in all seven domains drops by 12.57 points (8.36, 17.76) after accounting for their

Table 2
Participants' Ratings of Switching Intentions, Moral Conviction, and Acceptability of Maximization

	Switching intentions			Acceptability of maximization			Moral conviction		
Domain	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Advertisement	101	1.81	1.03	101	3.43	0.93	101	2.26	1.09
Agent	101	2.11	1.20	101	3.12	0.89	101	2.22	1.16
Plan	101	2.94	1.24	100	2.79	0.89	101	2.81	1.22
Premium	101	2.57	1.03	101	3.05	0.91	102	3.02	1.04
Prescription	101	3.38	1.23	101	2.48	0.98	101	3.63	0.96
Recommendation	101	2.68	1.16	101	2.84	0.90	101	2.87	1.04
Renewal	101	1.86	1.00	101	3.19	0.93	101	2.27	1.11

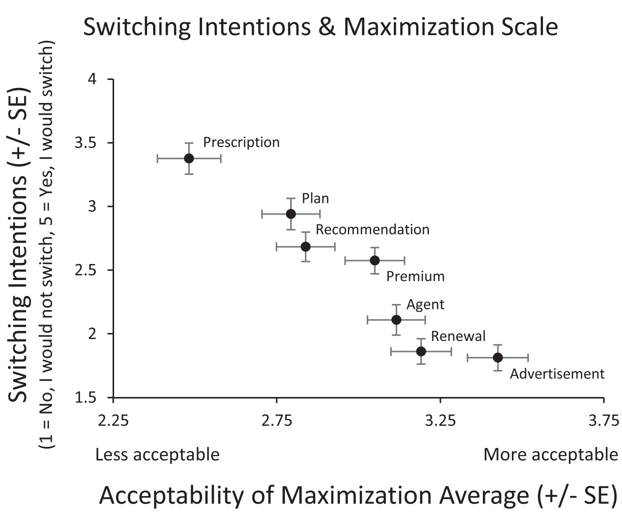

Figure 1. Switching intentions & acceptability ratings. *Note*. Participants rated their intention to switch insurance companies (higher numbers represent stronger intentions) and the appropriateness of using maximization for seven different types of decisions. Higher numbers on the maximization scale represent greater acceptability of maximization. This scatterplot shows the average rating for each of the seven types of decisions. Error bars show 1 standard error above/below sample means.

ratings of moral conviction. Completing this same analysis, but controlling for participants' rated acceptability of maximization in all regressions, results in a smaller 7.20 point drop in the *F*-value.[1] Thus, the drop in *F*-value associated with accounting for moral conviction shrunk by 5.37 points after controlling for participants' rated acceptability of maximization. A bootstrapped confidence interval of this reduction excludes zero (3.19, 8.20). This significant reduction shows that some of the variance in participants' moral conviction ratings that can account for their switching intentions is also accounted for by participants' ratings of the acceptability of maximization. This result is consistent with the notion that consumers are less open to algorithms making decisions in morally relevant domains because they feel that maximization is less acceptable in those domains.

Discussion

Study 2 finds that consumers express stronger objections to algorithmic decision makers when they feel that the decision in question is more morally relevant. These results suggest that marketers may face backlash when consumers learn that their company has adopted an algorithm for a morally relevant tradeoff. The results of Study 2 also suggest that consumers' resistance to algorithms in morally relevant domains stems from feelings that maximization is an inappropriate strategy in morally relevant tradeoffs. Thus, when consumers object to an organization using an algorithm to make a morally relevant tradeoff, they may do so because they believe that the organization is making this decision in a consequentialist manner. In Studies 3A & 3B, we use a different paradigm to examine these relationships with different paradigms and measures.

Studies 3A & 3B

We ran two similar studies designed to test whether there is a negative relationship between the moral relevance of a decision and consumers' openness to using an algorithm to make that decision, as proposed by hypotheses H2a and H2b. In

After running a regression of participants' switching intentions on decision domain dummies and the acceptability of maximization scale with no constant, the F-value on the hypothesis test that the domain dummies were equal was 17.50, $F(6, 698) = 17.50$, $p < .001$. After adding moral conviction to this regression, the F-value dropped 7.20 points, $F(6, 697) = 10.30$, $p < .001$; bootstrapped confidence interval (4.13, 11.41).

both studies, participants read about six different types of tradeoffs where resources are allocated to consumers, expressed their preference for a human versus an algorithm allocator, and rated their feelings of moral conviction for each type of tradeoff. In Study 3b, participants also rated their perceived importance of each type of decision in order to see whether participants' feelings of moral conviction predict their preferences between human and algorithmic decision makers after accounting for the perceived importance of the tradeoff.

Sample & Procedure

We preregistered that we would recruit 200 and 400 participants from Amazon Mechanical Turk for Studies 3a and 3b, respectively. For Study 3a, 237 participants clicked on the study link, 19 participants failed an attention check and were not allowed to begin the study, and 201 participants completed all dependent measures. For Study 3b, 469 participants clicked on the study link, 38 participants failed an attention check and were not allowed to begin the study, and 405 participants completed all dependent measures. These samples were 53% and 45% male, and 37 and 36 years old on average, respectively.

In both studies, participants who passed the attention check learned that they would read about 6 different types of allocation decisions. For each decision, one of two groups of people would be selected to receive a service or asset. Each of the groups had an average age of 20, 40, or 60 years old, consisted of 2, 4, or 6 people, and were 0%, 50%, or 100% female. We told participants to imagine that all groups were exactly the same except for these three characteristics. Participants learned that they would read about each type of decision and make two hypothetical decisions of each type, and then express their preferences and opinions regarding each type of decision.

Next, participants learned about each of the six types of decisions in randomized order. The decisions were as follows: *Coupons*—which of two groups should be selected to receive coupons for local stores, *Electricity Repair*—which of two groups should receive priority to have power company workers restore electricity in their home after a severe storm, *Flyer*—which of two groups should be selected to receive a flyer advertising a local store, *Store Greeting*—which of two groups should be greeted when entering a store at the same time, *Restaurant Seating*—which of two groups should receive the last available table at a restaurant before it closes for the day, and *Flood Rescue*—which of two groups should be rescued by a helicopter during a severe flood. For each type of decision, participants read a short description of the tradeoff on the first page, made one hypothetical allocation decision on the second page, and made a second hypothetical decision on the third page. For each hypothetical decision, the two groups that participants considered were randomly selected without replacement from a pool of 12 sets of two groups (see Supplement 2 in the Supplemental Materials document for a full description of the groups). The order of the group attributes (age, number, and % female) was randomized for each decision.

After learning about all six types of allocation decisions, participants read that they would express their preferences for having a human (another person) or an algorithm make each of the six types of decisions in real life. Participants read that both the other person and the algorithm are equally accurate and perform equally well. Next, participants expressed their preference for using human versus algorithmic decision makers for each type of decision in randomized order on a seven-point scale, where 1 = "I would only consider the person and never consider the algorithm" and 7 = "I would only consider the algorithm and never consider the person."

Next, participants expressed their feelings of moral conviction for each type of decision using the same questions as Study 2. These questions had a Cronbach's alpha ≥0.89 in each decision domain in each study. Additionally, participants in Study 3b completed scales measuring their perceived importance of each type of decision in randomized order. For each type of decision, participants in Study 3b completed three five-point scales that asked: "How costly is it to choose the incorrect group when deciding which group should [domain description]?", "How important is the decision of which group should [domain description]?", and "How high are the stakes when deciding which group should [domain description]?". We average these three ratings to create a measure of decision importance. Finally, participants reported their age, sex, and level of education.

Results

Moral conviction. Consistent with Hypothesis 2a, when participants found a decision domain to have more moral content, they expressed stronger preferences for human decision makers. We ran an OLS regression of participants' preferences between

Table 3
Participants' Ratings of Preference for the Algorithm, Moral Conviction, and Importance of the Decision

		Algorithm preference			Moral conviction			Decision importance		
Domain	*Study*	*N*	*M*	*SD*	*N*	*M*	*SD*	*N*	*M*	*SD*
Advertisement	3A	201	4.39	1.68	201	2.04	1.17	—	—	—
	3B	411	4.41	1.61	408	1.88	1.05	405	1.79	0.84
Agent	3A	201	3.47	1.77	201	3.27	1.10	—	—	—
	3B	411	3.45	1.80	410	3.18	1.12	407	3.21	0.89
Plan	3A	201	3.09	1.91	201	3.79	1.14	—	—	—
	3B	411	3.18	1.98	409	3.69	1.10	405	4.41	0.79
Premium	3A	201	4.41	1.59	201	1.95	1.13	—	—	—
	3B	411	4.51	1.55	410	1.77	1.01	405	1.70	0.84
Prescription	3A	201	3.82	1.72	201	2.36	1.16	—	—	—
	3B	411	3.72	1.65	409	2.16	1.09	406	1.62	0.79
Recommendation	3A	201	3.71	1.70	201	2.50	1.19	—	—	—
	3B	411	3.83	1.66	409	2.38	1.10	405	2.10	0.86

humans and algorithms on their moral conviction ratings with dummies to control for the decision domain and standard errors clustered by a participant ID. We found that within a decision domain, participants' moral conviction for the decision in question was negatively (positively) related to their preference for the algorithm (human) in Study 3a, $b = -0.23$, $t(201) = -2.98$, $p = .003$, and Study 3b, $b = -0.25$, $t(410) = -4.54$, $p < .001$. This is consistent with the notion that consumers are less willing to use an algorithmic decision maker for a specific decision when they feel that the decision in question has more moral content.

Consistent with Hypothesis 2b, participants expressed a stronger preference for human decision makers in more morally relevant decision domains. In the aggregate data describing the average ratings for each decision domain, participants' average ratings of moral conviction and preferences between human and algorithmic decision makers across the six types of decisions were significantly correlated in both Study 3a, $r(4) = -.95$, $p = .004$, and Study 3b, $r(4) = -.92$, $p = .009$, (see Table 3 & Figure 2). Thus, the aggregate data are consistent with the hypothesis that consumers express a stronger preference for human decision makers in more morally relevant domains.

Participants' hesitance to use algorithms in morally relevant domains also comes through in the individual data. We calculated the correlation between each individual participants' ratings of human/algorithm preference and moral conviction across the six different types of decisions. We found that 74.51% (114/153) and 72.05% (232/322) of participants expressed a negative correlation between their feelings of moral conviction and their preference for the algorithmic decision maker in Studies 3a and 3b, respectively. Wilcoxon signed-rank tests show that the median of participants' correlations, $-.55$ in Study 3a and -0.50 in Study 3b, was significantly less than zero in Study 3a, $z = -6.93$, $p < .001$, and Study 3b, $z = -9.06$, $p < .001$.[2]

Additionally, participants' ratings of moral conviction mediate the effect of the assigned decision domain on their preferences for human versus algorithmic decision makers. After running a regression of participants' preferences for human versus algorithmic decision makers on decision domain dummies with clustered standard errors and no constant, the *F*-value on the hypothesis test that the domain dummies were equal was 16.91 in Study 3a, $F(5, 200) = 16.91$, $p < .001$, and 35.69 in Study 3b, $F(5, 410) = 35.69$, $p < .001$. After adding participants' moral conviction ratings to these regressions, the *F*-value dropped 11.11 points in Study 3a and dropped 19.95 points in Study 3b. Bootstrapped confidence intervals of these drops in *F*-value exclude zero for both Study 3a (6.40, 17.61) and Study 3b (16.22, 32.57). This mediation analysis remains significant when controlling for the decision importance variable in Study 3b (3.41, 9.13). These findings are consistent with the notion that

These analyses excluded 48 participants in Study 3a and 87 participants in Study 3b because there was no variance in their responses to the human/algorithm preference questions or/and the moral conviction questions, and thus, the correlation calculation returned an error. The results do not meaningfully change when these participants are assumed to have a correlation of zero in Study 3a, z = −6.97, p < .001, and Study 3b, z = −9.03, p < .001.

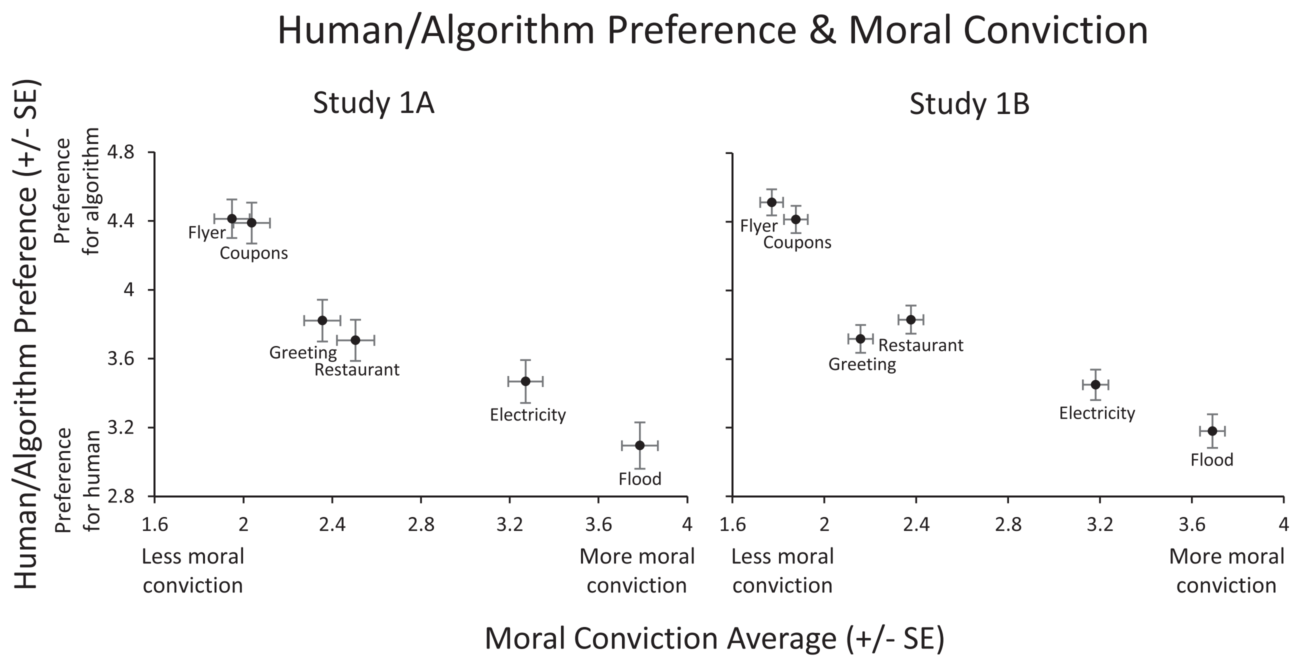

Figure 2. Preference for an algorithm & moral conviction. *Note*. Participants rated their preferences for a human (lower numbers) versus an algorithmic decision maker (higher numbers) and their feelings of moral conviction for six different types of decisions. Higher numbers on the moral conviction scale represent greater moral conviction. These scatterplots show the average rating for each of the six types of decisions. Error bars show 1 standard error above/below sample means.

consumers' tolerance for an algorithmic decision maker is at least partially determined by the moral relevance of the decision domain in question.

Decision importance. The results of Study 3b suggest that participants' feelings of moral conviction relate to their openness to using algorithms even after accounting for the importance of the decision. The three items measuring decision importance had a Cronbach's alpha of 0.96, and we combined them to create a scale of decision importance. We ran an OLS regression of participants' human versus algorithm preference for each decision on their ratings of moral conviction and decision importance and clustered standard errors by a participant ID. We found that participants' ratings of moral conviction were a significant predictor of their human versus algorithm preference, $b = -0.30$, $t(407) = -5.11$, $p < .001$, and that participants' ratings of decision importance were a marginally significant predictor of their human versus algorithm preference, $b = .09$, $t(407) = -1.86$, $p = .06$. A post-estimation test revealed that the coefficient of the moral conviction variable was larger than that of the decision importance variable, $F(1, 407) = 4.67$, $p = .031$, consistent with the notion that moral conviction may be better predictor of human versus algorithm preferences than decision importance.

Notably, these results hold when controlling for the decision domain. After adding six dummies representing the decision domain that participants were considering to the regression, moral conviction remained a significant predictor of human versus algorithm preferences, $b = -0.25$, $t(407) = -4.15$, $p < .001$, and decision importance became a non-significant predictor of human/algorithm preference, $b = .01$, $t(407) = .20$, $p = .84$. Further, the coefficient of the moral conviction variable remained significantly larger than that of the decision importance variable, $F(1, 407) = 6.34$, $p = .012$. Consistent with Hypothesis 2a, these results suggest that within a decision domain, participants who feel more moral conviction for a decision are less open to using an algorithmic decision even after controlling for their perceived importance of the decision.

Discussion

The results of Studies 3a and 3b support Hypothesis 2a—within a decision domain, those participants who felt that the decision was more morally relevant expressed a stronger preference for a human rather than an algorithmic decision maker. Further, this relationship persisted after controlling participants' perceived importance of the decision. Additionally, the results were consistent with

Hypothesis 2b—participants expressed a stronger preference for human decision makers in domains with more moral relevance, and participants' moral conviction ratings mediated the effect of the decision domain on their preference between a human and algorithmic decision maker. Along with the results of Study 2, this evidence suggests that consumers are much less tolerant of algorithmic decision makers in morally relevant domains, and that consumers may spurn organizations that use algorithms in morally relevant domains as a result.

Study 4

We designed Study 4 to further examine the link between consumers' concerns about maximization as a decision strategy in morally relevant tradeoffs and their intolerance for algorithmic decision makers making those tradeoffs. We ran a 2x2 study where we manipulated whether a human or algorithm would make a morally relevant decision (human, algorithm) and whether that decision-making entity would be trained to use maximization as a decision strategy (maximization training, no maximization training). If, as we have proposed, consumers' intolerance for algorithms in morally relevant domains is driven by their concerns about maximization as a decision strategy, then they should also express intolerance for a human decision maker who is trained to use maximization as a decision strategy. However, in line with Hypothesis 4, we would expect this training to have less of an impact on the evaluation of algorithmic decision makers, because consumers expect that algorithms are likely to use maximization even without explicit training (see Study 1). This is a high-powered replication of Study S1 in the supplemental materials (see Supplement 3 for details).

Sample & Procedure

We preregistered that we would recruit 1600 participants from Amazon Mechanical Turk. After posting the study, 1,707 participants clicked on the study link, 61 participants failed an attention check and were not allowed to begin the study, and 1,605 participants completed all dependent measures. This sample was 52% male and 40 years old, on average.

Participants who passed the attention check learned that they would read about a decision that a health insurance company makes and then answer relevant questions. Participants also read that there was a two-question comprehension check at the end of the survey and that they would get a $0.25 bonus if they answered both questions correctly. Next, all participants read about the prescription decision used in Studies 1 and 2. We chose this decision because participants expressed the most moral conviction for this domain in Study 2.

Next, participants were asked to imagine that the employee who used to decide which prescriptions were covered got promoted to a new job and that the company needs to choose how to make these decisions in the future. Participants assigned to the maximization condition read a passage explaining that the company will rely on maximization to make the prescription decisions. This passage told participants that the company's goal was, "maximizing specific concrete measures of outcomes when deciding whether or not to cover a medicine. These decisions will be made only by weighing measurable costs and benefits derived from data." Those not assigned to the maximization condition did not see this screen.

Next, participants assigned to the algorithm condition read that the company has decided to use a new algorithm to make the prescription decisions moving forward while those assigned to the human condition read that a new employee would make these decisions. Participants assigned to the maximization condition read that the new decision maker (i.e., the algorithm or human employee) would be trained to follow the maximization procedures described in the previous paragraph. After learning this information, participants were presented with a summary of the scenario up until this point. On the same page as the summary, they were asked to imagine that they were a customer of the insurance company and to report their switching intentions using the scale from Study 2.

Next, participants completed two comprehension check questions. These questions asked which decision the participant had read about (with four options, the correct answer being whether or not an insurance company will cover a medicine) and what prompted a change at the insurance company (with four options, the correct answer being that an employee got promoted). Participants then reported their age, sex, and highest completed level of education to complete the study.

Results

Consistent with Hypothesis 4, participants expressed a stronger negative reaction to a human employee receiving training to use maximization for a morally relevant decision than an algorithm (see Table 4 & Figure 3). We ran an OLS regression of participants' switching intentions on contrast coded variables representing whether they were assigned to the human (−1) versus the algorithm (1) condition, the maximization training (1) or no training (−1) condition, and the interaction between these two variables while excluding participants who failed either one of the comprehension check questions. The interaction term in this regression was negative and significant, $b = -0.10$, $t(1{,}496) = -3.44$, $p = .001$, consistent with the notion that training the human employee to use maximization caused a larger increase in switching intentions than training the algorithm to use maximization. When the 105 participants who failed one or both of the comprehension check questions are included in this analysis, this interaction term remains significant and the coefficient does not meaningfully change, $b = -0.09$, $t(1{,}614) = -3.47$, $p = .001$. These results suggest that people's objections to algorithms making morally relevant decisions may stem from their intolerance for maximization, the presumed way that algorithms approach decisions.

The coefficients of the main effects in this regression are also consistent with our hypothesizing. In line with Hypothesis 2, the coefficient on the variable representing the human versus algorithm condition revealed that participants expressed stronger switching intentions when they were assigned to read that an algorithm (versus a human) would make the morally relevant decision, $b = 0.31$, $t(1{,}496) = 10.95$, $p < .001$. In line with Hypothesis 3, the coefficient of the indicator representing training to use maximization was significant and positive, $b = 0.18$, $t(1{,}496) = 6.55$, $p < .001$, indicating that participants reported stronger switching intentions when they learned that the company would use maximization for a morally relevant decision. When the 105 participants who failed one or both of the comprehension check questions are included in this analysis, the coefficients of the algorithm, $b = 0.29$, $t(1{,}614) = 10.88$, $p < .001$, and maximization training terms, $b = 0.18$, $t(1{,}614) = 6.54$, $p < .001$, do not meaningfully change.

Table 4
Participants' Ratings of Switching Intentions

Condition	*N*	*M*	*SD*
Human employee, No Training	392	2.10	0.95
Human Employee, Trained to maximize	385	2.66	1.09
Algorithm, No Training	370	2.91	1.12
Algorithm, Trained to maximize	353	3.08	1.16

The values exclude participants who failed at least one comprehension check question.

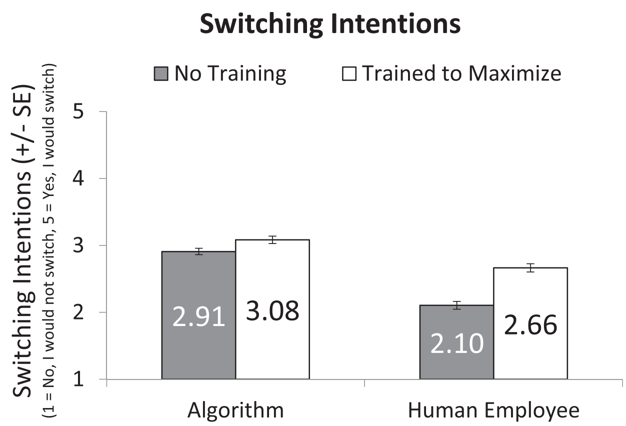

Figure 3. Responses to maximization training for humans versus algorithms. *Note.* We manipulated whether participants read that a human versus algorithm would make a morally relevant tradeoff at an insurance company, and whether that entity would be trained to make decisions using maximization. Participants rated their intention to switch insurance companies (higher numbers represent stronger intentions). The values represent the average response in each condition excluding participants who failed at least one comprehension check question. Error bars show 1 standard error above/below sample means.

Discussion

Study 4 finds that consumers may find maximization to be an intolerable decision strategy even when it is carried out by human employees. However, participants did not find a human employee who was trained to use maximization to be as intolerable as an algorithmic decision maker. This suggests that companies that need to employ a decision maker to work toward a maximization goal, like maximizing profit, may still benefit from choosing a human employee for this task instead of an algorithm.

General Discussion

The current studies suggest that consumers object to companies using algorithms to make morally relevant tradeoffs because they believe that algorithms will use maximization and find this decision process to be objectionable in morally relevant domains. The findings of Study 1 suggest that consumers may assume that algorithms are more likely to use the process of maximization to make a wide variety of consumer relevant decisions. Thus, when an organization employs an algorithm instead of a human employee to make a decision, consumers are likely to assume that the organization will approach that decision in a more consequentialist manner.

Studies 2 and 3 explore the consequences of using algorithms, which consumers believe to use maximization, for morally relevant decisions. For one, consumers who feel that a decision domain is morally relevant are more likely to object to the use of algorithms to make that decision, and spurn organizations that do so. This finding suggests that individual differences in consumers' perceptions of what is morally relevant can be an important factor that helps to explain why some consumers object to specific algorithms like self-driving cars, while others do not. Relatedly, the results of Studies 2 and 3 suggest that companies are more likely to face backlash for using algorithms to make decisions that are seen as being more morally relevant in general. Additionally, Study 2 finds that consumers object to algorithms in morally relevant domains specifically because they object to the use of maximization in these domains.

Study 4 finds that consumers also object to maximization as a decision strategy when it is performed by a human employee, consistent with the notion that maximization is what makes consumers object to algorithmic decision makers in morally relevant domains. However, participants still expressed stronger objections to an algorithm using maximization than a human employee using maximization, possibly because consumers believe that human decision makers will still consider other factors, like deontological rules, even after being trained to use maximization. This result suggests that when managers need their organization to maximize some outcome measure (e.g., profit, engagement, impressions), this practice may be more tolerable when it is carried out by human employees.

Managerial Implications

These findings have many implications for marketers. When consumers feel that a decision domain is relatively amoral, they may be more open to using an algorithm to make that decision or patronizing an organization that uses an algorithm to make that decision. Indeed, some work suggests that people may be open to using algorithms in various amoral domains (Logg, Minson, & Moore, 2019), in which other factors may determine consumers' openness to relying on algorithms. For example, consumers' willingness to rely on an algorithm may depend on: whether or not they have seen it make a mistake (Dietvorst, Simmons, & Massey, 2015), whether the task at hand is subjective (Castelo et al., 2019), whether they have an identity motivation for their consumption decision (Leung et al., 2018), and whether outcomes in the domain are subject to randomness (Dietvorst & Bharti, 2020). However, in the vast majority of studies that have found that people are open to using algorithms under certain conditions, those algorithms have been operating in relatively amoral domains, like making amoral predictions (see Dietvorst & Bharti, 2020; Dietvorst et al., 2015; Logg et al., 2019).

On the other hand, our work suggests that consumers may not be open to using algorithms in morally relevant domains. In morally relevant domains, the consequentialist approach that algorithms use to make decisions may be generally unacceptable to consumers as a decision strategy. Thus, we would expect that regardless of the other considerations that researchers have investigated regarding the adoption of automation, consumers may object to companies using algorithms in morally relevant domains. Anecdotally, there is a dearth of examples of consumers embracing an algorithm for a morally relevant decision in the literature.

It is unclear how to make algorithms acceptable in morally relevant domains. It is possible that developing an algorithm using a process other than maximization, and making this process explicit to consumers, could make for an acceptable algorithm. However, it is not clear what algorithmic decision process consumers would find to be acceptable. Alternatively, an algorithm could carry out maximization but place constraints on the maximization so that it does not violate any moral rules; however, it is not clear that this is possible given the multitude of potentially relevant moral rules and the diversity of people's moral values (see Nozick, 1974).

It is worth noting that heterogeneity in consumers' preferences could also be an obstacle for marketing algorithms. Given the heterogeneity in participants' moral conviction for different decision domains, it is probable that a non-trivial portion of consumers may see a decision as being morally relevant and be unwilling to use an algorithm even when the majority of consumers see the decision to be relatively amoral. Given the potential backlash that companies could face in these cases, this issue is worth serious consideration on the part of marketing managers.

Theoretical Implications

The findings in this paper speak to consumers' adoption of algorithmic decision makers across various domains. Most notably, this work suggests that whenever consumers find maximization to be an objectionable approach for decision-making, they may find the use of algorithms to be unacceptable. For example, it is possible that consumers prefer investment decisions that have moral undertones to be made by humans instead of algorithms (see Niszczota & Kaszás, 2020) specifically because they expect that algorithms will use maximization to make these decisions and find this approach to be objectionable. Additionally, consumers may prefer not to use algorithms in the domain of medicine not only because they expect that algorithms cannot account for humans' unique characteristics (see Longoni et al., 2019), but also because they find maximization to be an inappropriate way to weigh the harms and benefits that can result from medical decisions.

Further, these results speak to consumers' theories about the way that algorithms operate. Specifically, consumers seem to believe that algorithms operate by combining information to maximize (or minimize) some predetermined objective. We believe that this presumed decision process will not only make algorithms objectionable in morally relevant domains, but may also steer people away from algorithmic decision makers in domains where this process seems incompatible with their goals. For example, consumers who have identity motives may resist automation (see Leung et al., 2018) because their goal of expressing themselves (e.g., through their cooking, driving, fashion, etc.) is incompatible with products that automatically and independently pursue a predetermined objective. People may prefer to use human judgment instead of algorithms for subjective tasks (see Castelo et al., 2019) because they feel that the dogged pursuit of one specific goal is incompatible with problems that lack a universal answer (e.g., choosing art for one's home). Finally, consumers may resist self-driving cars, because programming cars with a predetermined objective (e.g., prioritize the safety of drivers, prioritize the safety of pedestrians unless they break a rule, protect the largest group, etc.) is necessarily incompatible with a desire for these cars to "protect me no matter the circumstances" (see Bonnefon et al., 2016).

This work also serves as an additional example of people's objections to relying on consequentialist decision strategies when making morally relevant tradeoffs (see Baron & Spranca, 1997; Bartels et al., 2016; Everett et al., 2016; Goldsmith et al., 2018; Tetlock, 2002; Tetlock et al., 2000). As explained in the theoretical development section, maximization is essentially a consequentialist decision strategy because of its sole focus on obtaining optimal outcomes regardless of any process considerations. Like the past work cited above, the current studies' results suggest that people may find consequentialist strategies to be unacceptable for humans to carry out in morally relevant tradeoffs. Further, it seems that people find consequentialist decision processes to be even more objectionable when they are carried out by algorithms; perhaps because of algorithms' presumed unwavering pursuit of their objective coupled with a complete ignorance of other considerations like deontological rules.

Boundaries of Intolerance for Algorithms in Morally Relevant Decisions

We do not expect that people will necessarily object to using algorithms to make any decision that has moral relevance. Specifically, algorithms may be acceptable for very straightforward morally relevant decisions that do not require a tradeoff. As other scholars have pointed out, many morally relevant decisions are straightforward and have an obvious acceptable answer (Kamm, 2008). These decisions tend to be non-controversial because they are not tradeoffs, which means that all desirable outcomes can be achieved simultaneously. As a result, most any normative theory (Deontology, Utilitarianism, etc.) or decision maker would come to the same conclusion. For example: Should a man be given a requested medical intervention that poses no cost or risk to others? Should a company make a costless modification to a product that will save three lives? Should a community center be built that faces no objections from anyone in the community and will pay for itself? We believe that algorithms that pick obviously acceptable answers for these types of decisions are likely to be tolerable (as long as consumers are not worried about an algorithm picking the wrong option).

However, if these straightforward decisions are transformed into tradeoffs, we would expect algorithms to be intolerable to consumers. For example: Should a man be given a requested medical intervention which requires a scarce resource that could be used to save others? Should a company make a modification to a product that will save three lives but cost 15 million dollars? Should a community center be built that will only benefit a fraction of those who will pay for it and faces objections? Because these decisions are tradeoffs, they require the decision maker to weigh harms and benefits against each other. The empirical evidence in this paper suggests that people will find maximization to be an unacceptable decision strategy to balance these harms and benefits, and object to algorithmic decision makers as a result.

Further, people could object to algorithms more strongly the more complicated morally relevant tradeoffs get. In the current studies, participants faced tradeoffs with a limited number of considerations; however, real-life tradeoffs are often far more complex. We anticipate that algorithms may be even more intolerable for these more complex morally relevant tradeoffs because: (a) the pursuit of any one predetermined objective may be even less likely to align with any given consumer's preferences, (b) algorithms may be more likely to inadvertently violate a moral rule, and (c) people prefer decision makers to show thoughtful consideration when making complex moral decisions (Landy, Herzog, & Bartels, 2021).

Limitations & Future Directions

The studies in this paper do have limitations and leave open questions. First, we did not present a study where we manipulated how moral a tradeoff is while holding everything else constant (our reading of the literature suggests that, despite researchers' best attempts, no one else has either). We attempted to run studies of this nature; however, our attempts either failed to manipulate moral conviction or manipulated other factors in addition to moral conviction. Thus, one could be concerned that we have not isolated the effect that moral conviction has on human/algorithm preference. We took a number of steps to address this concern. We used multiple stimuli within (e.g., Studies 3a, 3b, and 2 contained 6 or 7 decision domains) and across (e.g., Studies 2 and 3 used completely different scenarios) studies, suggesting that the results are not a function of strategic or fortunate choices of stimuli. In Study 3b, we controlled for what we considered to be the most obvious confounding variable (i.e., the importance of the decision) and found that moral conviction still predicts human/algorithm preferences. Finally, we report mediation analyses that are consistent with the notion that it is the differences in participants' perceived moral conviction across domains that drives the differences in their human/algorithm preferences across domains. Although we believe that these measures may help to alleviate this concern to some degree, we also understand that they do not eliminate it.

Second, it is possible that the results of our studies could have turned out differently if we had made different choices when designing and conducting the studies. For example, there could be (a) ways of describing algorithms that makes them more tolerable for moral tradeoffs, (b) moral domains where consumers do not exhibit intolerance of algorithms, or (c) populations for which the results would turn out differently, to name a few possibilities. Future research could investigate the generality of these patterns across these factors.

Third, our results do not suggest a way for organizations to get consumers to adopt or accept algorithms for moral tradeoffs. There are many potential avenues, like (a) modifying the way that algorithms make decisions, (b) reframing morally relevant tradeoffs to make them seem more amoral, and (c) educating consumers, among others. The general issue of non-adoption of algorithms for important, consequential, moral tradeoffs is a crucial issue for future research to address.

Conclusion

In sum, consumers object to maximization as a decision process for making morally relevant tradeoffs. Further, because people assume that algorithms are based on maximization, they are intolerant of organizations using algorithms to make morally relevant tradeoffs. These results help to explain why consumers may object to algorithms operating in some domains (e.g., medicine and criminal justice) while accepting them in others (e.g., entertainment and navigation), and suggest that companies should be conscious of potential backlash when applying algorithms to or publicizing algorithms in morally relevant domains.

References

Alexander, S., & Ruderman, M. (1987). The role of procedural and distributive justice in organizational behavior. *Social Justice Research*, *1*, 177–198.

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*, 1–16.

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381–417.

Bartels, D., Bauman, C., Cushman, F., Pizarro, D., & McGraw, A. P. (2016). Moral judgment and decision making. In G. Keren, & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 478–515). Chichester, UK: Wiley.

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science*, *5*, 187–202.

Bhattacharjee, A., Berman, J. Z., & Reed, A. (2013). Tip of the hat, wag of the finger: How moral decoupling enables consumers to admire and admonish. *Journal of Consumer Research*, *39*, 1167–1184.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*, 1573–1576.

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

Campbell, M. C., & Winterich, K. P. (2018). A framework for the consumer psychology of morality in the marketplace. *Journal of Consumer Psychology*, *28*, 167–179.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*, 809–825.

Davis, N. (1993). Contemporary deontology. In P. Singer (Ed.), *A companion to ethics* (pp. 205–218). Oxford, UK: Blackwell.

Deghani, M., Atran, S., Iliev, R., Sachdeva, S., Medin, D. L., & Ginges, J. (2010). Sacred values and conflict over Iran's nuclear program. *Judgment and Decision Making*, *5*, 540–546.

Deghani, M., Iliev, R., Sachdeva, S., Atran, S., Ginges, J., & Medin, D. L. (2009). Emerging sacred values: Iran's nuclear program. *Judgment and Decision Making*, *4*, 930–933.

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science*, *31*, 1302–1314.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*, 114–126.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*, 1155–1170.

Dunn, J. (2017, June 27). Pandora has gained 5 million users in the past 3 years, while Spotify has gained 100 million. *Business Insider*. Retrieved from http://www.businessinsider.com/pandora-vs-spotify-total-subscribers-chart-2017-6

Ehrich, K. R., & Irwin, J. R. (2005). Willful ignorance in the request for product attribute information. *Journal of Marketing Research*, *42*, 266–277.

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*, 772–787.

Goldsmith, K., Roux, C., & Ma, J. (2018). When seeking the best brings out the worst in consumers: Understanding the relationship between a maximizing mindset and immoral behavior. *Journal of Consumer Psychology*, *28*, 293–309.

Irwin, J. R., & Naylor, R. W. (2009). Ethical decisions and response mode compatibility: Weighting of ethical attributes in consideration sets formed by excluding versus including product alternatives. *Journal of Marketing Research*, *46*, 234–246.

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

Kagan, S. (1998). *Normative ethics*. Boulder, CO: Westview Press.

Kamm, F. M. (2000). Nonconsequentialism. In H. LaFollette (Ed.), *The Blackwell guide to ethical theory* (pp. 205–227). Oxford, UK: Blackwell.

Kamm, F. M. (2008). *Intricate ethics: Rights, responsibilities, and permissible harm*. Oxford, UK: Oxford University Press.

Kim, H. Y., & McGill, A. L. (2018). Minions for the rich? Financial status changes how consumers see products with anthropomorphic features. *Journal of Consumer Research*, *45*, 429–450.

Konovsky, M. A. (2000). Understanding procedural justice and its impact on business organizations. *Journal of Management*, *26*, 489–511.

Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable algorithms. *University of Pennsylvania Law Review*, *165*, 633.

Landy, J. F., Herzog, N. R., & Bartels, D. M. (2021). *Moral thoughtfulness: Thinking carefully about complex moral problems is a virtue*. Working Paper.

Leinert, P. (2018, January 28). Most Americans wary of self-driving cars: Reuters/Ipsos poll. *Reuters*. Retrieved from https://www.reuters.com/article/us-autos-selfdriving-usa-poll/most-americans-wary-of-self-driving-cars-reuters-ipsos-poll-idUSKBN1FI034

Leung, E., Paolacci, G., & Puntoni, S. (2018). Man versus machine: Resisting automation in identity-based consumer behavior. *Journal of Marketing Research*, *55*, 818–831.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. New York, NY: Plenum Press.

Lior, R. (2014). *Data mining with decision trees: Theory and applications* (Vol. *81*). Hackensack, NJ: World Scientific.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*, 629–650.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, *38*, 937–947.

McGraw, P. A., Schwartz, J. A., & Tetlock, P. E. (2011). From the commercial to the communal: Reframing taboo tradeoffs in religious and pharmaceutical marketing. *Journal of Consumer Research*, *39*, 157–173.

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, *56*, 535–556.

Niszczota, P., & Kaszás, D. (2020). Robo-investment aversion. *PLoS One*, *15*, e0239277.

Nozick, R. (1974). *Anarchy, state, and utopia*. New York, NY: Basic Books.

Perez, S. (2016, July 21). Google Maps is turning its over a billion users into editors. *Tech Crunch*. Retrieved from https://techcrunch.com/2016/07/21/google-maps-is-turning-its-over-a-billion-users-into-editors/

Reczek, R. W., Irwin, J. R., Zane, D. M., & Ehrich, K. R. (2017). That's not how I remember it: Willfully ignorant memory for ethical product attribute information. *Journal of Consumer Research*, *45*, 185–207.

Rhoades, L., & Eisenberger, R. (2002). Perceived organizational support: A review of the literature. *Journal of Applied Psychology*, *87*, 698.

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In *Recommender systems handbook* (pp. 1–35). Boston, MA: Springer.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, *27*, 135–153.

Shaffer, V. A., Probst, C. A., Merkle, E. C., Arkes, H. R., & Medow, M. A. (2013). Why do patients derogate physicians who use a computer-based diagnostic support system? *Medical Decision Making*, *33*, 108–118.

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, *4*, 267–281.

Skitka, L. J., Bauman, C. W., & Lytle, B. L. (2009). Limits on legitimacy: Moral and religious convictions as constraints on deference to authority. *Journal of Personality and Social Psychology*, *97*, 567.

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review*, *109*, 451.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo tradeoffs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*, 853.

Thottam, I. (n.d.). 10 Online dating statistics you should know. *Eharmony*. Retrieved from https://www.eharmony.com/online-dating-statistics/

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Tyler, T. R. (1988). What is procedural justice?: Criteria used by citizens to assess the fairness of legal procedures. *Law and Society Review*, *22*, 103–135.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, *32*, 403–414.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Supplementary Material: Modern Algorithm.