Title
Consumer Responses to Algorithmic Decisions

Author
Demirdag, Bilge Ipek

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Consumer Responses to Algorithmic Decisions

A dissertation submitted in partial satisfaction of the

requirements for the degree of Doctor of Philosophy

in Management

by

Bilge Ipek Demirdag

2022

ABSTRACT OF THE DISSERTATION

Consumer Responses to Algorithmic Decisions

by

Bilge Ipek Demirdag

Doctor of Philosophy in Management

University of California, Los Angeles, 2022

Professor Suzanne B. Shu, Chair

It has been long acknowledged that computational prediction procedures may yield more accurate predictions than human judges. Nevertheless, people are often algorithm averse, that is, they are less willing to rely on algorithms than humans in tasks such as forecasting. Previous research on algorithm aversion has largely examined algorithmic *forecasts* or *recommenders* and not algorithmic *decisions*. This dissertation explores an uninvestigated facet of algorithm aversion: consumer attitudes and behavior regarding decisions that algorithms make on their behalf. Consumer responses to technology that performs its operations without any human involvement and is autonomous has been recognized as an important construct that needs to be studied in consumer research. As algorithms are increasingly becoming autonomous decision-makers, it is crucial to study how consumers perceive and react to algorithmic decisions.

Chapter 1 encompasses five pre-registered studies (combined $N = 2{,}535$) conducted across diverse digital domains. It highlights consumers' divergent conceptualizations of human and algorithmic decisions and suggests that consumers perceive algorithms as *black boxes*, whereas they perceive humans as more transparent. The lower satisfaction with algorithmic decisions is accounted for by lower trust in algorithms, which results from consumers' perception that the algorithm's decision is less transparent relative to human decisions. I find that increased *input explainability* (i.e., the consumer's ability to access relevant input information regarding a particular decision) is an effective intervention to increase transparency and trust, leading to higher consumer satisfaction with algorithmic decisions.

Chapter 2 investigates consumers' perceptions of bias in algorithms and humans. The findings of four studies (combined $N = 3{,}121$) demonstrate a "bias tolerance" phenomenon, i.e., people acknowledge but disregard human bias and trust human decisions more than algorithmic ones. Algorithmic decisions are perceived as less biased, but paradoxically as less trustworthy and satisfactory than human decisions. This is because the negative effect of human (vs. algorithm) bias on satisfaction is less than the positive effect emotionality has on satisfaction. I find boundary conditions for bias tolerance in tasks (material purchases and data handling) where human emotionality and bias is impertinent.

Across two chapters, this dissertation contributes substantively and theoretically to our comprehension of how consumers' divergent conceptualization of human and algorithmic decision processes influences their responses to those decisions. As algorithms are increasingly becoming autonomous decision-makers, understanding how consumers perceive and react to algorithmic decisions can allow us to determine methods, such as input explainability, for a more satisfactory consumer-algorithm interaction.

The dissertation of Bilge Ipek Demirdag is approved.

Stephen Spiller

Keith Chen

Franklin Shaddy

Suzanne B. Shu, Committee Chair

University of California, Los Angeles

2022

DEDICATION

In loving memory of my grandparents:

Yesare Serdar, Dr. Salih F. Serdar, Bilge Yazıcılar, Muhlis Demirdağ

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Chapter 1: Insights into the Black Box: Process Transparency of Algorithmic Decisions Drives Consumer Satisfaction in the Digital World

A version of this chapter is invited for 3rd round revision: Demirdag, I. & Shu, S. B. Insights into the Black Box: Process Transparency of Algorithmic Decisions Drives Consumer Satisfaction in the Digital World. *Journal of Consumer Research.* I designed the studies in collaboration with S. B. Shu. I collected the data, conducted the analyses, and prepared the manuscript. S. B. Shu provided advice and edited the manuscript.

Chapter 2: Bias Tolerance: When Human Bias, but not Algorithmic Bias, is Disregarded

A version of this chapter is being prepared for publication: Demirdag, I. & Shu, S. B. Bias Tolerance: When Human Bias, but not Algorithmic Bias, is Disregarded. I designed the studies in collaboration with S. B. Shu. I collected the data, conducted the analyses, and prepared the manuscript. S. B. Shu provided advice and edited the manuscript.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my family for their unconditional love and support that allowed me to finish my PhD. I am incredibly grateful to my mom, Ayşım Demirdağ, my dad, Reha Demirdağ, and my sister, İdil Demirdağ, for all the uplifting motivation that they have given me throughout my life. Without you, none of this achievement would have been possible. All through my PhD, there has never been a day where I did not feel your support. I know how remarkably lucky I am to have you as my family. This is as much your accomplishment as it is mine.

Secondly, I would like to express my very great appreciation to my advisor, Suzanne Shu, for her positivity, support, and encouragement throughout my PhD. Through her guidance, I felt like I could conquer any and all research questions I may have, which allowed me to think outside the box. She provided me with an indispensable skillset to be a great independent researcher, which will be immensely valuable as I proceed in my research career.

Thirdly, I would like to thank my committee members Stephen Spiller, Keith Chen, and Franklin Shaddy. They have offered invaluable suggestions on my research papers since their inception. My papers are exceedingly better due to their insightful feedback that they have given me in the last several years.

I would also like to thank all my friends. I am truly fortunate that I got to start my PhD with exceptional friends: Joey Reiff, Julia Levine, David Zimmerman, Jon Bogard, Sherry He, and Pedro Makhoul. I would not have enjoyed this process nearly as much if it were not for all these wonderful friends. I would also like to thank my best friend of 14 years, Tuğçe Şık, for her never-ending love and support. Despite not having lived in the same city for a long time, we only grew closer, and our friendship kept me going in challenging times.

Last but not least, I would like to thank my little research assistant, my cat, Carl. He has been with me from start to finish in all my research projects. He always kept me company and was in the same room as I was as I wrote every single word in this dissertation. His unlimited love motivated me as I went through the ups and downs of this process.

CURRICULUM VITA

IPEK DEMIRDAG

EDUCATION

UNIVERSITY OF MICHIGAN April 2017
Master of Science in Psychology (Cognition and Cognitive Neuroscience Concentration)

UNIVERSITY OF MICHIGAN April 2016
Bachelor of Science in Biopsychology, Cognition, and Neuroscience (with Distinction and Honors)
Minor in Applied Statistics

MANUSCRIPTS UNDER REVIEW

Demirdag, Ipek and Suzanne B. Shu, "Insights into the Black Box: Process Transparency Drives Consumer Satisfaction in the Digital World," invited for 3rd round review at the *Journal of Consumer Research*.

SELECTED WORK IN PROGRESS

Demirdag, Ipek and Suzanne B. Shu, "Bias Tolerance: When Consumers Tolerate Human Bias, but not Algorithmic Bias," in preparation for the *Journal of Marketing Research*.

Demirdag, Ipek and Suzanne B. Shu, "Ownership Where Ownership Does Not Exist: Psychological Ownership in Sharing Economies," in preparation for the *Journal of Consumer Psychology*.

Demirdag, Ipek and Franklin Shaddy, "Psychology of Technology Failure," data collection in progress.

Demirdag, Ipek and Suzanne B. Shu, "The Effects of Consumer-Algorithm Goal Alignment," data collection in progress.

INVITED TALKS AND SELECTED CONFERENCE PRESENTATIONS

Consumer Responses to Algorithmic Decisions (2021)

- Invited talk presented at Yale University, Dartmouth College, Cornell University, University of Texas-Austin, University of South Carolina, Tulane University, San Diego State University

Ownership Where Ownership Does Not Exist: Psychological Ownership in Sharing Economies (2020)

- Conference talk presented at the 2020 Society for Consumer Psychology Conference and the 2020 Association for Consumer Research Conference

Bias Neglect: When Consumers Neglect Human Bias, but not Algorithmic Bias (2020 and 2021)

- Conference talk presented at the 2020 Association for Consumer Research Conference, the 2021 Society for Consumer Psychology Conference, and the 2021 University of Houston Doctoral Symposium

Insights into the Black Box: Process Transparency Drives Consumer Satisfaction in the Digital World (2019 and 2020)

- Conference talk presented at the 2019 Society for Judgment and Decision Making Conference and the 2020 Society for Consumer Psychology Conference

TEACHING AND MENTORSHIP EXPERIENCE

Teaching Assistant, MGMTFE 411: Marketing Management **Spring 2021**
Professor Franklin Shaddy
UCLA Anderson School of Management

Teaching Assistant, MGMTEX 411: Marketing Management **Spring 2020**
Professor Suzanne Shu
UCLA Anderson School of Management

Teaching Assistant, MGMT 411: Marketing Management **Fall 2018 and Fall 2019**
Professors Hal Hershfield and Franklin Shaddy
UCLA Anderson School of Management

Research Team Leader **Fall 2016-Spring 2017**
In a laboratory environment, supervise and teach undergraduate students various aspects of psychological research, such as literature review, collecting and analyzing data, and producing written reports.
Judgment and Decision Lab, University of Michigan

SELECTED HONORS, AWARDS, AND GRANTS

UCLA Dissertation Year Fellowship ($20,000 in recognition of contributions to UCLA's intellectual community) **2021**
Psychology of Technology Institute Dissertation Award Honorable Mention **2021**
AMA Sheth Foundation Doctoral Consortium Fellow **2020**
Morrison Center Research Funding, UCLA Anderson School of Management **2020**
UCLA Anderson Summer Doctoral Fellowship **2018-2022**
UCLA Anderson Doctoral Fellowship **2017-2022**
Rackham Graduate Student Research Grant **2016**
Tanner Memorial Award (top research project), University of Michigan **2016**
James B. Angell Scholar Award, University of Michigan **2015-2016**
LSA Honors Program, University of Michigan **2015-2016**
Conference Travel Award, University of Michigan **2015**
LSA Honors Grant for Research and Travel, University of Michigan **2015**
University Honors, University of Michigan **2013-2015**

CHAPTER 1: INSIGHTS INTO THE BLACK BOX: PROCESS TRANSPARENCY OF ALGORITHMIC DECISIONS DRIVES CONSUMER SATISFACTION IN THE DIGITAL WORLD

ABSTRACT

Algorithms have moved from being prediction tools into being decision makers for a variety of consumer-relevant outcomes. Although algorithms are often more accurate than humans, consumers may be wary of algorithmic decisions. This article explores *perceived process transparency* (i.e., the level at which a decision process is perceived to be understood) as a novel driver of possible algorithm aversion. Five pre-registered studies (combined $N = 2{,}535$) conducted across a variety of digital domains highlight consumers' divergent conceptualizations of human and algorithmic decisions and suggest that consumers perceive algorithms as *black boxes* with opaque decision processes, whereas they perceive themselves and other humans as more process transparent. This lower perceived process transparency leads to lower trust in algorithms and negative effects on satisfaction for algorithmic decisions. Increasing *input explainability* (i.e., the consumer's ability to know relevant input information regarding a particular decision) is revealed as an effective intervention to increase both process transparency and trust, leading to higher consumer satisfaction with decisions made in the digital world.

Algorithmic decisions are increasingly prevalent. Banks use algorithms to determine whether our credit card application is approved or not. Courts use algorithms for bail and sentencing decisions. Uber picks our car and driver for a given ride. YouTube decides what we should watch next. As with many digital platforms, the decisions made in these domains are usually not binding, but these algorithms are gradually transforming from recommenders to being primary decision-makers. For instance, Amazon has patented a system for anticipatory package shipping, in which it deploys an algorithm to predict what customers are going to buy before they even place an order and preemptively sends those products to them (Joel R. Spiegel et al. 2013). Similarly, the Netflix CEO Reed Hastings has said, “One day we hope to get so good at suggestions that we are able to show you exactly the right film or TV show for your mood when you turn on Netflix,” (The Economist 2016). As we continuously provide more data to these digital platforms, algorithms’ predictive accuracies are increasing to the extent that they may know us better than we know ourselves. This paves the way for bypassing the personalized recommendation stage and directly making decisions for consumers.

Previous research on algorithms reveals a complex account for how people react to algorithms. On one account, people are algorithm averse. That is, they reject the use of algorithms in favor of human decisions, especially when algorithms are observed making mistakes and for tasks that are viewed as more subjective (Castelo, Bos, and Lehmann 2019; Dietvorst, Simmons, and Massey 2015). On another account, people appreciate algorithms, meaning that they prefer algorithms over humans for tasks related to geopolitical forecasts, song popularity, stock market performance, and navigation (Castelo et al. 2019; Logg, Minson, and Moore 2019). Castelo, Bos, and Lehmann (2019) summarize individuals’ reactions to algorithmic advice in domains including health (Promberger and Baron 2006), humor (Yeomans et al. 2019), and finance (Önkal et al. 2009); much of this literature documents that people prefer or put more weight on the advice that comes from other humans relative to advice from algorithms. Reasons for preferring human recommendations can include beliefs that humans are more fair, flexible, ethical, accurate, and better to account for unique circumstances. When algorithms are preferred, level of expertise can moderate the effect, as experts rely more on their own judgments than those of an algorithm (Logg et al. 2019). We expand on this prior work about when algorithms are liked or disliked by also considering how perceived transparency of algorithms can affect their trustworthiness and acceptability.

We propose that underlying the concern many people have about algorithms is that they are seen as *black boxes*. That is, they lack both perceived and actual process transparency and are overly convoluted for humans to understand. When consumers observe outputs of algorithms, the internal functioning of how they use information to reach decisions may not be easily comprehensible. More concretely, we propose that a driver of consumer aversion towards algorithmic decisions is *perceived process transparency*. By perceived process transparency, we indicate the amount by which a decision process is perceived to be understood. To a consumer, their own internal decision processes and another human's decision processes are more likely to be judged as transparent than an algorithms' decision processes. As a consequence of algorithms being perceived as inscrutable black boxes, consumers may find algorithmic decisions less trustworthy and satisfactory than decisions that humans make. On the other hand, if consumers understood how an algorithm made its decision, they may trust it more and regard its decision as more acceptable. Based on the importance of process transparency for decision acceptance, this article tests an intervention to increase algorithmic transparency and trust, and therefore, decision satisfaction. This intervention is *input explainability*, defined as consumer's ability to know relevant input information regarding a particular decision.

Across five pre-registered studies[1], we reveal that process transparency is a crucial factor for consumer trust and satisfaction in the digital space. We find that consumers perceive algorithms as less transparent than humans, which leads to lower trust in and satisfaction with algorithmic decisions than human decisions. Furthermore, we explore input explainability as an effective intervention to increase trust and satisfaction with choices in digital space.

TRUST IN ALGORITHMS: REALITY VERSUS PERCEPTIONS

Algorithms are data-based, machine-driven prediction procedures that vary in complexity. Simple algorithms such as linear regressions, logistic regressions or decision trees can be interpreted by inspecting their parameters. For instance, star ratings, reviews, and helpfulness ratings can be used as inputs in a multiple linear regression to predict customer sentiment for a product. In contrast, complex models such as deep neural networks are black box models that are computationally elaborate and also highly performant. For example, an online retailer may employ deep neural networks trained on history of interactions to fully understand what the customer wants and to hyper-personalize the experience such as by providing personalized offers in real time. These various algorithms that are often used in forecasting, advice, recommendations, or decision-making may or may not be worthy of trust.

[1] For all experiments, we have reported all measures, conditions, and data exclusions. You can find our data, pre-registrations, and original generating files of our studies at https://osf.io/hcx2q/?view_only=6fcd4ef908a648f79ec3e77a6afcb6ea.

Trust has been acknowledged as an important construct for human-automation relationships that increases adoption of automation (Lee and See 2004; Moray, Inagaki, and Itoh 2000; Muir 1987; Muir and Moray 1996). Trust is willingness to be vulnerable to another party's actions, contingent on the positive expectations about that party's behavior (Mayer, Davis, and Schoorman 1995; Rousseau et al. 1998). Due to the perceived risk of automation, whether and when people are willing to be vulnerable to automated decisions is critical.

Trust is a multidimensional construct that can indicate that another will act according to the principles of competence, benevolence, integrity, or predictability (McKnight and Chervany 2001). Our paper concentrates on the first component, competence, i.e., having the capability to accomplish what needs to be accomplished. This is based on past research showing that judgments of ability and performance determine trust in automation and subsequently acceptance of that technology (Choi and Ji 2015; Gefen, Karahanna, and Straub 2003; Lee and Moray 1992). As trust in automation is largely based on competence (Muir and Moray 1996), our study of trust will center on consumers' perceptions of an algorithm's ability to perform the task at hand. Competence-based trust can vary depending on consumers' judgments of how accurate an algorithm performs a task. Nonetheless, a distinction needs to be made between the reality and perceptions of trustworthiness, as they do not always align perfectly.

The reality is that not all algorithms are worthy of being trusted in making decisions, as their outputs can be inaccurate and augment human biases by various means such as training data and data preparation. For instance, algorithms that assess a criminal defendant's likelihood of becoming a recidivist and algorithms used for medical predictions of costs and illness have been found to display racial bias (Angwin et al. 2016; Obermeyer et al. 2019). Discriminatory algorithms can be found in areas such as credit scoring as well (Rice and Swesnik 2013).

Algorithms used in price discrimination, where different consumers are charged different prices for the same product (Hannak et al. 2014), can also be concerning to customers. For instance, it was discovered that Orbitz, an online travel website, was showing more expensive hotel options to Mac users than PC users when they discovered that Mac users were spending about 30% more per night on hotel rooms (Mattioli 2012), and The Princeton Review was using ZIP codes to charge different groups of customers different prices for the same SAT tutoring course (Angwin, Mattu, and Larson 2015).

People are against the use of algorithms that may include bias. According to a Pew Research Center survey (2018), the majority of respondents find automated decision systems such as resume screening of job applicants, video analysis of job interviews, and personal finance scores unacceptable (57%, 67%, and 68% respectively). The same survey states that 56% of Americans consider criminal risk assessments unacceptable. In Wisconsin, a man who was sentenced to prison based on an algorithmic prediction of recidivism appealed the decision, citing that the secret nature of this algorithm prevents him from "challenging the accuracy and scientific validity of the risk assessment" (Supreme Court of Wisconsin 2016). In cases such as these, consumers' distrust in black box algorithms is more than justified, as these algorithms promote existing societal biases.

Notwithstanding these serious concerns about possibly inaccurate and biased algorithms, there are many instances where algorithms can improve human decision-making and yet people still perceive them as untrustworthy. It has been long acknowledged that computational prediction procedures may yield more accurate predictions than human judges in a variety of settings such as medicine, mental health, organizations, and education (Camerer 1981; Dawes 1979; Dawes, Faust, and Meehl 1989; Grove et al. 2000; Meehl 1954). And despite the failure of some recidivism prediction algorithms, other machine learning algorithms are found to be reducing bias in sentencing decisions compared to human judges (Kleinberg et al. 2017). The use of such an algorithm can reduce the jail population by 42% without elevating crime rates, while also shrinking the population of minorities in jail. Résumé-screening algorithms can also improve human bias by providing more opportunities to non-traditional candidates who are more productive employees, despite not having a degree from an elite university, job referrals, or prior experience (Cowgill 2020). Algorithms can even help corporations perform better by choosing the directors that will be popular with the shareholders (Erel et al. 2017).

All these highly performant algorithms are trained on past, biased data. Therefore, they can be biased too, but they are less biased and more accurate (i.e., less noisy) than humans (Kahneman, Sibony, and Sunstein 2021). Human experts can even be outperformed by random linear models that use information consistently, unlike humans (Dawes and Corrigan 1974; Yu and Kuncel 2020). Therefore, consumer decisions can benefit from algorithmic methods that enhance consistency and reduce noise. Even so, people are often algorithm averse, and less willing to rely on data-based, machine-driven prediction procedures than humans in tasks such as forecasting and decision-making (Arkes, Shaffer, and Medow 2007; Dietvorst et al. 2015; Eastwood, Snook, and Luther 2012; Highhouse 2008; Önkal et al. 2009; Promberger and Baron 2006). This is particularly true when the algorithms make mistakes (Dietvorst et al. 2015) and for tasks that are perceived as subjective (vs. objective) (Castelo et al. 2019). Although algorithms could improve their judgments and decisions, consumers are cautious of using them.

Despite the evidence that consumers are wary of algorithmic decisions, there are also cases where consumers may actually prefer to rely on algorithms over human decisions. For instance, people exhibit algorithm appreciation when forecasting geopolitical and business events, song popularity, and romantic attraction (Logg et al. 2019). Additionally, people trust algorithms more than humans in predicting stock market performance, weather forecast, and giving directions (Castelo et al. 2019). People are also more confident in an algorithmic forecast than themselves when they can even slightly modify the algorithmic forecast (Dietvorst, Simmons, and Massey 2018).

In sum, algorithms may be realistically trustworthy although they are not always seen as perceptually trustworthy. Consumers have varying levels of algorithm acceptance across different types of tasks. In this article, we focus on perceived trustworthiness of algorithms from the consumer's perspective. Perceptions are useful for marketers to understand since they are pertinent for whether consumers are comfortable using algorithms. We study when and why consumers view algorithms as untrustworthy and its impact on satisfaction for consumer outcomes. On that account, we present a novel determinant of trust: process transparency.

THE ROLE OF PROCESS TRANSPARENCY IN DETERMINING TRUST

Decision recipients value process transparency and knowing how an outcome materializes. For instance, people are less likely to be aggrieved at adverse outcomes if they know that the procedures used to acquire those outcomes are fair (Brockner 2002; Brockner and Wiesenfeld 1996) and if people do not understand a recommendation process, they are less inclined to use that recommender system (Yeomans et al. 2019). When making their own decisions, consumers think that they have great introspection abilities and can report their cognitive processes accurately even though their actual processes are not fully transparent and can be influenced by stimuli outside their awareness (Nisbett and Wilson 1977; Wilson 2004).

Given that consumers' attitudes and behaviors are shaped by self-assessed knowledge, which is weakly associated with actual knowledge (Alba and Hutchinson 2000), it is important to measure consumers' self-assessed transparency of their own decision-making processes alongside their assessment of other humans' and algorithmic decision-making processes.

In our investigations, we differentiate between three different decision types: decisions made by the self, decisions made by another human, and decisions made by the algorithm. A comparison of only self choice and algorithmic choice would raise the question of whether the real difference is between the self and any other third party, regardless of if it is an algorithm or a human. Past research has documented a multitude of differences in responses to self decisions and other's decisions. People tend to discount other humans' advice and weigh their own judgments more (Harvey and Fischer 1997; Lim and O'Connor 1995; Yaniv and Kleinberger 2000), potentially because they can evaluate the reasons for their own judgments better than they can evaluate them for another person (Yaniv 2004), again suggesting the important role of perceived process transparency. When it comes to algorithms, however, the gap between human and non-human may lead to even less of an understanding of the underlying reasoning and processes.

We propose that these possible different conceptualizations of algorithmic decision processes and human decision processes is an important driver of algorithm acceptance or aversion. In particular, we suggest that consumers perceive algorithms as *black boxes*, themselves as *transparent boxes*, and other humans as *semitransparent boxes*. When consumers observe outputs of algorithms, the internal functioning of how they use information to reach decisions may not be as easily comprehensible as if they or another similar human were to make the same decision. In other words, people may judge other humans' decision processes as less transparent than their own but more transparent than an algorithms' decision processes. This would be consistent with the past research showing that people thinking they have more in common with human advisors than algorithmic advisors (Prahl and Van Swol 2017). As a result, despite the potential benefits of algorithmic decisions, consumers may opt for human choice and be uncomfortable with the idea of letting an algorithm choose on their behalf unless they believe that they understand the algorithm's decision process. We can consider a clothing choice context. Consumers may think that know the underlying processes for how they, as transparent boxes, picked a clothing combination but lack that knowledge for a stylist (a semitransparent box) and even more so for an algorithmic stylist (a black box). This may be why Stitch Fix, which is an online personal styling service, has created an "Algorithms Tour" explaining how they use data science and machine learning for personalization (Colson et al. 2021). To the extent that this Algorithms Tour enhances process transparency, consumers may enjoy the clothes that the algorithmic stylist picked.

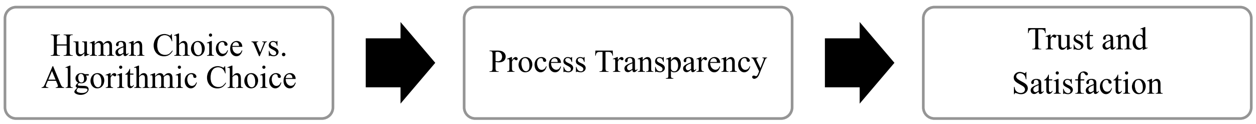
The implications of higher perceived process transparency in any decision process are higher levels of trust and satisfaction. As noted earlier, knowledge of decision procedures can lead to more satisfaction with adverse outcomes (Brockner 2002; Brockner and Wiesenfeld 1996). Process fairness promotes trust as well (Korsgaard, Schweiger, and Sapienza 1995; Sapienza and Korsgaard 1996). We propose that perceived process transparency, regardless of fairness concerns, can also directly influence trust. Accordingly, we study the hypothesis that consumers may prefer and trust human (self or other's) choice over algorithmic choice due to higher perceived process transparency of human (vs. algorithmic) decisions. We note that this does not apply to all algorithms – for example, consumers have widely adopted cruise control in cars and autopilot on planes despite not having full process transparency – but repeated interactions with these algorithms and immediate feedback on their accuracy has likely increased trust for these algorithms over time. We test this idea later in the paper by measuring trust for algorithms that are already well-liked and regularly used versus those that are not.

In sum, we theorize that if consumers do not feel like they understand the underlying mechanisms of decisions made in the digital space, they may not be comfortable with being vulnerable to an algorithmic decision-maker or believe in its ability to perform the decision task under consideration, and consequently be dissatisfied with the decision. In contrast, since consumers perceive their decision processes as more transparent, they are more satisfied with choosing themselves than letting a less transparent entity (i.e., another human or an algorithm) choose on their behalf. We further posit that consumers are more accepting of decisions that come from another human more than those that come from an algorithm due to the perception that algorithms are less process transparent than other humans. That is, a further implication of our account is that to the degree another human's choice (vs. algorithmic choice) is higher in perceived process transparency, it should yield higher trust and decision satisfaction. Putting these pieces together, we propose that for many domains where algorithms may replace human decision-makers, the level of perceived transparency of algorithmic (vs. human) decision processes affects consumers' trust and satisfaction with the decision-maker's choice (Figure 1). Accordingly, the following hypotheses are proposed:

H1a: Trust in and satisfaction for algorithmic choice is lower than that for human (self or other) choice.

H1b: The decreased trust and satisfaction from algorithmic (vs. human) choice is mediated by lower perceived process transparency for algorithms.

FIGURE 1 (CHAPTER 1). CONCEPTUAL MODEL FOR THE EFFECT OF HUMAN (VS. ALGORITHMIC) CHOICE



INPUT EXPLAINABILITY TO ENHANCE PROCESS TRANSPARENCY

While consumers would appreciate greater knowledge about an algorithm's process, this is not always possible. Instead, we propose an alternate path to enhance perceived process transparency: input explainability. By *input explainability*, we refer to consumers' knowledge of relevant input information regarding a particular decision. We hypothesize that provision of inputs to a decision will create insights regarding the decision process, even without explicit knowledge of the actual process. That is, consumers believe they can infer the underlying decision process by being cognizant of the decision inputs. Companies can thus enrich consumers' knowledge about how their black box algorithm operates simply by disclosing the inputs. We propose that input explainability interventions can thus enhance perceived process transparency necessary for trust and satisfaction.

This prediction is based on dividing up a decision system into three key parts: (1) the inputs (the *why* of the decision), (2) the process (the *how* of the decision), (3) the output (the *what* of the decision). For decisions made on their behalf, either by algorithms or by other humans, consumers may know the what (the output) without necessarily having knowledge of the why or how. We propose that knowledge about *why* a decision was made (i.e., explainable inputs) will allow inference about *how* a decision was made (i.e., transparent process) to generate the outcome, thus leading to greater satisfaction with the overall decision system.

A focus on explaining why a decision was made is valuable for two main reasons. The first reason is substantive. In our daily lives, we may need explanations for many reasons: (1) predicting when a similar event will occur in the future, (2) diagnosing why a system malfunctioned, (3) justifying an action, (4) blaming and penalizing the guilty party in a one-time event, and (5) enhancing the appreciation of an observation (Keil 2006). Thus, providing explanations enhances consumer experiences. For instance, corporations that offer explanations in which they accept responsibility for a problem receive the most positive customer reaction (Conlon and Murray 1996). Explanations accompanying forecasts are also important in determining trust in forecasting (Önkal, Gönül, and De Baets 2019). In addition to substantive explanations, providing placebic explanations that lack meaningful information can increase compliance with a request (Langer, Blank, and Chanowitz 1978). Just as we seek explanations from other people or we engage in explaining ourselves to others, we may expect and want algorithms to provide explanations for their decisions. Similar to the effects of placebic explanations, explaining algorithm inputs without greater detail on actual processes may be enough to increase trust.

Second, disclosing the *why* as opposed to the *how* is often more feasible and preferrable for companies. Companies may be more likely to implement a program where they reveal which pieces of information are used in order to arrive at a decision, rather than the inner workings of their proprietary algorithm which could be copied by a competitor. Even if companies are willing to disclose their source code, which might be incomprehensible for most consumers (Caruana et al. 2015; Lakkaraju et al. 2017), the true nature of black box algorithms may remain nontransparent if they continue to change as the machine learns (Rudin 2019). For such complex algorithms, the inputs are knowable, even when the process is not. Prior efforts to reveal a black box model's internals (the *how*) and mathematically explain its predictions have been found to not increase individuals' adherence to the model predictions (Poursabzi-Sangdeh et al. 2018). In these instances where the process explanations are highly technical, the explanatory power of inputs can still provide consumers with a perceived causal chain between the inputs and the outputs, without the explicit process information.

Input explainability can lead to perceived process transparency if the knowledge of inputs suggests simplistic explanations, as it has been shown that people prefer simple explanations (Lombrozo 2007). While knowledge of the inputs (the *why*) to a decision is not the same as full knowledge of the internal decision-making process (the *how*), providing information about the content of the inputs to an algorithm may help consumers construct possible interpretations of the decision process. As in the Howard-Sheth model of consumer behavior (Howard and Sheth 1969), input variables inform psychological constructs that influence consumer decision processes. By the same token, input explainability interventions that provide knowledge about the inputs for a decision can improve perceived process transparency and trust, and in turn increase algorithm acceptance.

We also note that the influence of input explainability applies not only to algorithmic decision-makers, but also to decisions made by humans (self or other). Although we postulate that algorithms are perceived as less transparent than humans, we also suggest that to the extent that human decision processes are perceived as nontransparent, a consumer will distrust another human's ability to choose and have decreased contentment with the outcome. This distrust could also be overcome by making decision processes appear more transparent via disclosing the inputs. Our studies investigate whether an input explainability intervention can increase perceived process transparency and, therefore trust, in decisions that both algorithms and other people make, consequently increasing decision satisfaction. Therefore, the following hypothesis is proposed:

H2a: Input explainability increases trust in decisions through enhanced perceived process transparency.

We additionally predict that input explainability will have a differential effect on humans and algorithms. This prediction builds on H1b, where we hypothesize that algorithmic decisions will be perceived as less process transparent and therefore less trustworthy than human decisions. On that account, if algorithms are less process transparent at baseline than humans, boosting input explainability of decisions should enhance trust more for algorithms than humans. Furthermore, diminishing input explainability should be more detrimental to human decisions than algorithmic decisions, since humans are already perceived as more process transparent by default. We thus predict that input explainability will be a more effective intervention for algorithmic (vs. human) decisions.

H2b: The effect of increased input explainability on trust is greater for decisions made by algorithms than for decisions made by humans.

CURRENT RESEARCH

Consumer responses to technology that performs its operations without any human involvement and is autonomous has been recognized as an important construct that needs to be studied in consumer research (Schmitt 2019). This article explores consumers' conceptualizations of human and algorithmic decision processes. Specifically, we propose that lower perceived process transparency explains the lower trust in and preference against algorithmic decisions relative to human decisions (H1a and H1b). Importantly, we test whether input explainability can increase process transparency and trust with the decisions made in the digital world (H2a), as well as whether this effect is stronger for algorithms than humans (H2b).

We test these predictions across five pre-registered studies. In study 1a, we test hypothesis 1a by inquiring into possible differences in perceived process transparency, trust, and satisfaction between self choice and algorithmic choice in a dating decision context. We also test hypothesis 1b where we investigate process transparency as an underlying mechanism that accounts for higher trust and satisfaction in self choice over algorithmic choice. In study 1b, we provide evidence for hypotheses 1a and 1b in a risky decision context, where higher process transparency of the self drives trust in and comfort with a gambling bet that had higher odds of losing than winning. Study 2 contributes to our understanding of process transparency by adding choice by other humans to our investigation of choice types (i.e., self choice and algorithmic choice). Study 3 builds on previous studies by examining if the role of process transparency for explaining trust is similarly substantial in situations where algorithmic decisions are more standard. Study 4, which tests hypotheses 2a and 2b, introduces an input explainability intervention to see if it can augment perceived process transparency and therefore trust, leading to increased satisfaction in a highly consequential context. All together, these five studies strive for a better insight into the effects of consumers' divergent conceptualizations of human and algorithmic decisions. We report the pertinent visual stimuli and additional analyses in the appendix.

STUDY 1A: PROCESS TRANSPARENCY AND TRUST IN SELF CHOICE VERSUS ALGORITHMIC CHOICE (DATING APP)

Study 1a tests H1a by probing whether consumers view algorithmic decision processes as less transparent than their own decision processes. Furthermore, it investigates whether consumers trust algorithmic decisions less than their own decisions. Per H1b, study 1a tests whether perceived process transparency accounts for consumers' level of trust in and satisfaction with a decision. If consumers view algorithmic (vs. self) decisions as less process transparent, then that might explain why they would view algorithmic decisions as less trustworthy. We test these hypotheses in a dating app context. Prior research has found that individuals preferred algorithms over humans when forecasting romantic attraction (Logg et al. 2019), but they preferred advice from humans over algorithms in a dating service (Castelo et al. 2019). Consistent with this literature, in a situation where either the consumers themselves or an algorithm could make a partner choice from a dating app, we predicted that consumers would predict trusting their own choice and finding the resultant date more satisfactory.

Method

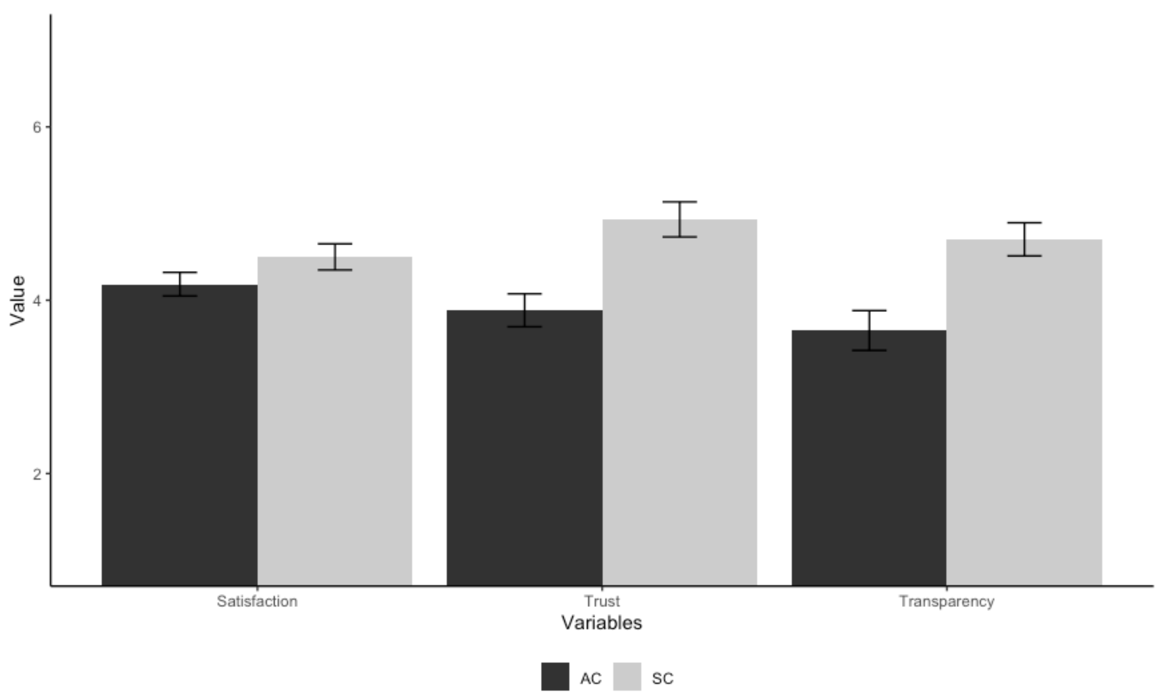
Participants. Two hundred and ninety-four (M_{age} = 22.9, SD = 3.9, 71.1% female) participants from a U.S. university's subject pool were recruited. They were randomly assigned to one of two conditions that varied in terms of choice type: Self Choice (SC), Algorithmic Choice (AC).

Materials and Procedure. Participants were asked to imagine that they were single and that they use a new dating app called "DateNow." Participants in the Self Choice condition were told that in this app, they get to choose who they will go on a date with from among 5 people each week. Participants in the Algorithmic Choice condition were told that in this app, the DateNow algorithm will choose who they will go on a date with from among 5 people each week. Next, participants indicated how willing they would be to go on a date with the person they or the algorithm picked from this app (1 = "Very unwilling", 7 = "Very willing"), how satisfied they would be with the person they or the algorithm picked (1 = "Very unsatisfied", 7 = "Very satisfied"), and the extent they trusted their or the algorithm's ability in deciding on the person they would like to go on a date with (1 = "Strongly distrust", 7 = "Strongly trust"). They then completed the perceived process transparency measure by indicating how well they felt they could understand the processes behind their or the algorithm's choice of which person to go on a date with on a 7-point scale (1 = "Little to no understanding", 4 = "Moderate understanding", 7 = "Detailed and deep understanding", adapted from Sloman and Rabb 2016). Finally, participants completed a measure for preference for autonomy (adapted from Botti, Orfali, and Iyengar 2009) as a control variable. Participants then rated how much they would like having (or not having) to make this decision of which person they would go on a date with from the app (1 = "Not at all", 7 = "Extremely"). As people have an innate psychological need for autonomy (Ryan and Deci 2000), which would be lacking in algorithmic choice, we wanted to ensure that the differences between self and algorithmic choice on our dependent variables were above and beyond people's desire for autonomy.

Results

Planned Comparisons. There was a significant effect of choice type on satisfaction, trust, and process transparency (Figure 2). Algorithmic (vs. self) choice participants expected to be less satisfied with their date choice (M_{SC} = 4.50, M_{AC} = 4.18, $F(1, 292) = 9.22$, $p = .003$, R^2 = .03), trusted the algorithm's date choice less (M_{SC} = 4.93, M_{AC} = 3.88, $F(1, 292) = 54.86$, $p <$ .001, $R^2 = .16$), and found the decision process less transparent (M_{SC} = 4.70, M_{AC} = 3.65, $F(1, 292) = 47.68$, $p < .001$, $R^2 = .14$). Although directionally AC participants were less willing (vs. SC) to go on a date, there was no significant difference on this measure (M_{SC} = 4.97, M_{AC} = 4.82, $F(1, 292) = 1.34$, $p = .249$, $R^2 = .005$).

FIGURE 2 (CHAPTER 1). STUDY 1: ALGORITHMIC (VS. SELF) CHOICE WAS LESS PROCESS TRANSPARENT, TRUSTWORTHY, AND SATISFACTORY

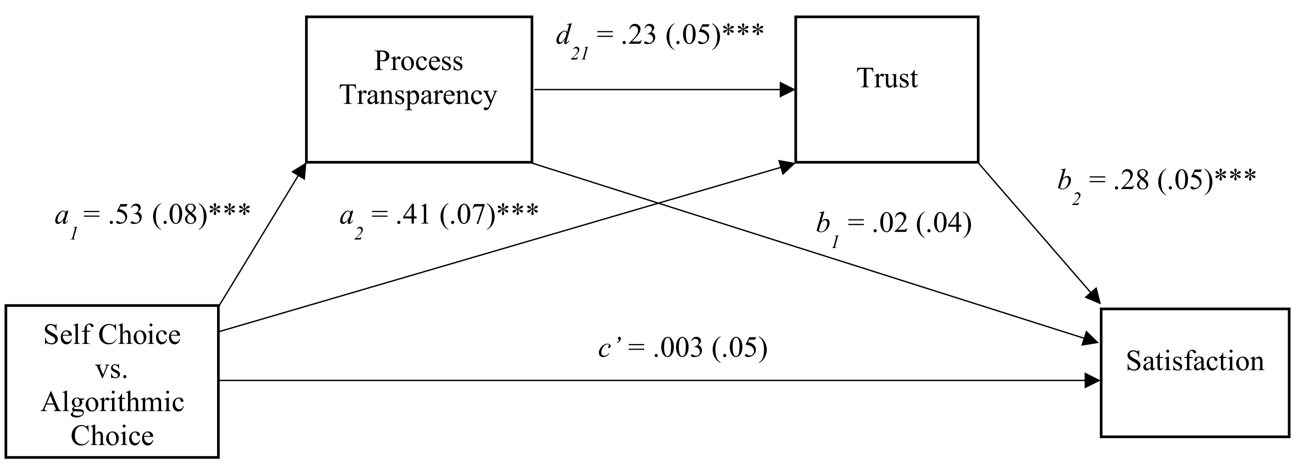


Note. Error bars represent ± standard error of the mean (SEM).

Analysis on the preference for autonomy control variable indicated that AC participants liked not having to make this date choice less than SC participants liked having to make this choice (M_{SC} = 5.45, M_{AC} = 3.88, $F(1, 298)$ = 38.73, p < .001, R^2 = .12). Consequently, we ran the linear regressions looking at the effect of choice type on our dependent variables while controlling for this preference for autonomy. Results showed that AC's (vs. SC's) lower satisfaction, trust, and process transparency were replicated even when controlling for preference for autonomy (see Appendix A, Section 1 for details).

Mediation Analysis. We conducted a serial mediation analysis using 10,000 bootstrapped samples to test our hypothesis that AC (vs. SC) leads to lower perceived process transparency, and therefore, lower trust, reducing satisfaction. We found a significant indirect-only mediation of perceived process transparency for the effect of choice type on decision trust and therefore satisfaction, indirect effect = .034, SE = .011, 95% CI = [.017, .062] (Figure 3, Model 6 of Hayes 2013).

FIGURE 3 (CHAPTER 1). STUDY 1A: ALGORITHMIC (VS. SELF) CHOICE WAS LESS PROCESS TRANSPARENT, THEREFORE LESS TRUSTWORTHY AND SATISFACTORY



Note. *** p < .001. Parameter estimates and standard errors (in parentheses) are indicated.

Discussion

The results of study 1a demonstrate that algorithmic choice was perceived as less process transparent, trustworthy, and satisfactory than self choice, even after controlling for preference for autonomy. Furthermore, we were able to establish perceived process transparency as a novel driver of algorithm acceptance. In particular, we found that perceived process transparency explains the relationship between choice type, trust and satisfaction. Study 1a was conducted in a college dating decision context, where the risk to participants in receiving a poor outcome from among a pool of other students may be seen as minimal. In study 1b, we investigate a possible downside of higher perceived process transparency for self choice versus algorithm choice by testing whether these measures can affect behavior in a riskier context, such as being comfortable with a gambling bet with negative expected value.

STUDY 1B: CAN PROCESS TRANSPARENCY ABOUT SELF CHOICE LEAD TO RISKY BEHAVIOR? (GAMBLING)

Study 1a provided initial evidence in support of H1a that consumers view algorithmic decision processes as less transparent and trustworthy than their own decision processes. Furthermore, it showed that lower perceived process transparency leads to consumers' lower trust in algorithms choosing for them. Study 1b was designed to show another possible downside of high process transparency about the self versus an algorithm: comfort with risky decisions, such as gambling decisions, that are made by the self versus an algorithm. To test the effects of risk, study 1b employs a roulette gambling scenario at a Las Vegas casino. The expected value of a $1 bet on red or black in a roulette spin is -$0.053. The odds against winning are $1\frac{1}{9}$ to 1.

Therefore, any bet on a color is inherently a risky decision. However, if individuals feel like they understand their own decision process to pick one of the colors to bet on, they might feel more comfortable betting on that color. In contrast, if an algorithm picks a color for them, they might feel less comfortable with the bet, since they would lack process transparency for how the algorithm made its color decision. Study 1b tests these predictions by demonstrating how consumers are more comfortable with risky behavior when they (vs. algorithms) decide because they believe they understand their decision process and, therefore, trust their ability to decide.

Method

Participants. We recruited two hundred and ninety-six Amazon Mechanical Turk (MTurk) participants (M_{age} = 35.3, SD = 11.1, 37.8% female). They were randomly assigned to one of two conditions: Self Choice, Algorithmic Choice.

Materials and Procedure. All participants were instructed to imagine that they were visiting Las Vegas and that they were playing roulette in their hotel casino. We told them that we were simulating a real roulette game in which they would see a spin of a fair roulette wheel and that they could earn a bonus based on their decisions (i.e., choices were incentive compatible).

Self Choice participants chose the color (red or black) to bet on, whereas Algorithmic Choice participants read that a roulette algorithm app on their phone told them to bet on one of the colors (participants were randomly shown either red or black as the app choice). Next, they wrote about how they or the roulette algorithm app decided on which color to bet on. Then, they rated process transparency by indicating how well they understand the process behind their or the algorithm's choice of that particular color on a 7-point scale (1 = "Little to no understanding", 4 = "Moderate understanding", 7 = "Detailed and deep understanding", adapted from Sloman and Rabb 2016). They rated the extent they trusted their or algorithm's ability in deciding on which color to bet on a 7-point scale (1 = "Strongly distrust", 7 = "Strongly trust"). Then, they were told that they had $25 to gamble and were asked to enter in the maximum amount (between $1 and $25) they would bet on the chosen color. To make this bet incentive compatible, we told them that they could earn up to $.50 depending on how much they have won or lost after playing the roulette wheel. How well they did was determined by their bet amount and whether their chosen color matched the color that the ball fell onto after the spin. Specifically, we awarded a bonus of $.50 for participants who bet anywhere between $13 and $25 and won the roulette. We paid a bonus of $.25 for participants who bet between $1 and $12 and won the roulette.

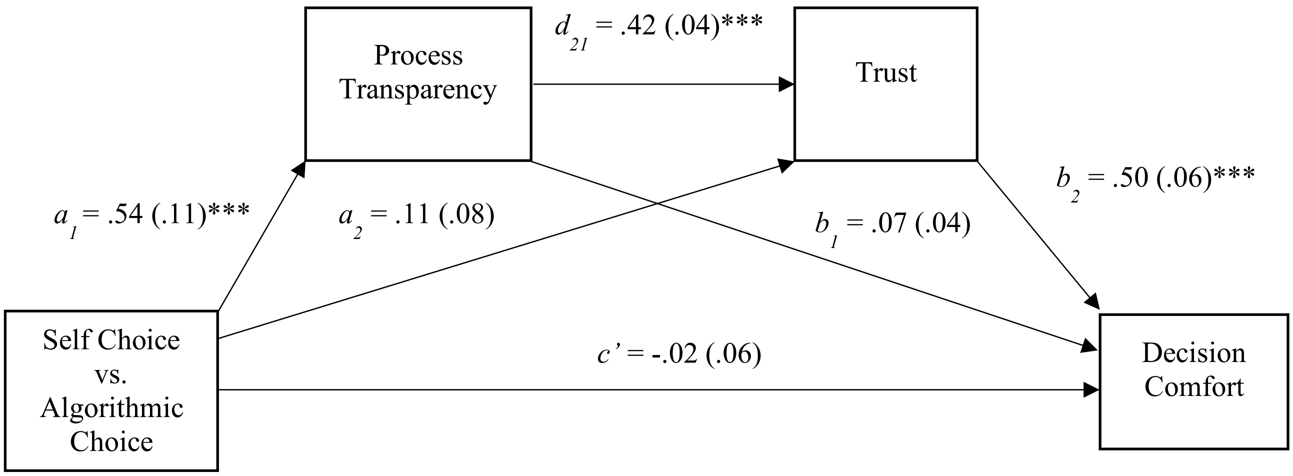
After placing the bet but before seeing the outcome, participants indicated the extent that they were comfortable with the bet they placed on the chosen color (1 = "Very uncomfortable", 7 = "Very comfortable"). Additionally, we included measures for uniqueness neglect as a control variable. Past research has shown that one of the drivers of algorithm aversion is uniqueness neglect, which is the concern that algorithms are less capable to consider consumers' unique characteristics and circumstances than humans (Longoni, Bonezzi, and Morewedge 2019). By including this measure, we aimed to demonstrate that low perceived process transparency exerts an independent effect beyond uniqueness neglect and that algorithm aversion is not driven by consumers' concerns that the algorithmic decision does not account for their uniqueness. Participants rated the extent to which they agreed with 3 statements as they pertain to the choice scenario they read: "I (This algorithm) would not recognize the uniqueness of my tastes," "I (This algorithm) would not consider my unique circumstances," and "I (This algorithm) would not tailor the decision to my unique case" (1 = "Strongly disagree", 7 = "Strongly agree", adapted from Longoni, Bonezzi, and Morewedge 2019). These three items were averaged to derive a uniqueness neglect index ($\alpha = .92$). Finally, the roulette wheel was spun, and the winning color was displayed. Participants' bonus payment was allocated according to their bet.

Results

There was a significant effect of choice type on process transparency, trust, and comfort. Self (vs. algorithmic) choice participants perceived their decision as more process transparent ($M_{SC} = 4.82$, $M_{AC} = 3.74$, $F(1, 294) = 25.44$, $p < .001$, $R^2 = .08$) and trustworthy ($M_{SC} = 5.15$, $M_{AC} = 4.47$, $F(1, 294) = 14.68$, $p < .001$, $R^2 = .05$). Furthermore, SC (vs. AC) participants were more comfortable with the gamble they placed on the chosen color ($M_{SC} = 5.43$, $M_{AC} = 5.05$, $F(1, 294) = 5.91$, $p = .016$, $R^2 = .02$). These results were replicated when controlling for uniqueness neglect (see Appendix A, Section 2 for details). Our exploratory analysis on the amount of the bet placed before the our comfort DV showed no difference between SC and AC ($M_{SC} = 15.47$, $M_{AC} = 15.85$, $F(1, 294) = .162$, $p = .688$).

We ran a serial mediation analysis with a bootstrapping procedure using 10,000 samples which uncovered an indirect-only mediation for the effect of the SC (vs. AC) leading to higher decision comfort through process transparency and trust, indirect effect = .114, SE = .027, 95% CI = [.068, .178] (Figure 4, Model 6 of Hayes 2013). This serial mediation indicated that the effect of algorithmic choice on decision comfort was driven by lower trust in the algorithm, which resulted from people's belief that the algorithm's decision was less process transparent than one's own decisions.

FIGURE 4 (CHAPTER 1). STUDY 1B: SELF (VS. ALGORITHMIC) CHOICE IS MORE PROCESS TRANSPARENT AND, HENCE, MORE TRUSTWORTHY, LEADING TO HIGHER DECISION COMFORT



Note. *** $p < .001$. Parameter estimates and standard errors (in parentheses) are indicated. Indirect effect (a_1*d_{21}*b_2) was .114, SE = .027, $p < .001$, 95% CI = [.068, .178].

Discussion

The results of study 1b demonstrate that perceived process transparency explains the relationship between choice type and trust. In particular, algorithmic choice was perceived as less process transparent and, hence, less trustworthy than self choice. Furthermore, we were able to establish perceived process transparency as a novel driver of algorithm acceptance, rather than uniqueness neglect which did not account for the difference in trust levels between self and algorithmic choice.

Study 1b encapsulates a possible downside of higher perceived process transparency of the self: individuals are more likely to be comfortable with risky decisions, such as a roulette bet where the expected value of winning is negative, because they view their decision processes as more transparent and therefore more trustworthy than algorithmic decisions. In other words, the additional trust and comfort that comes from a perception of high process transparency can sometimes lead to non-optimal decisions, such as a chance of losing money. In these circumstances, more use of algorithmic decision systems may help individuals reduce their risky behaviors.

STUDY 2: WHAT ABOUT ANOTHER HUMANS' CHOICE? (CAR ACCIDENT)

Studies 1a and 1b showed that consumers view algorithmic decision processes as less transparent and trustworthy than their own decision processes, per H1a. Moreover, they demonstrated a mechanism by which lower perceived process transparency explains consumers' lower trust in algorithmic (vs. self) choice, per H1b. Study 2 makes two main contributions. First, we advance our study of choice types by incorporating another human's choice. It could be that the effects we observed were merely due to the differences in self and a third party, not specifically due to the differences between the self and the algorithm. Second, we test our hypotheses in a decision domain with high moral implications: a car accident.

The investigation of transparency and trust for another human's choice is an important comparison to algorithmic choice. Does it matter that it is an algorithm making a choice versus just an entity outside oneself? Determining whether consumers consider algorithms to be different from all humans (self and others) or just from the self is important for establishing a broader perspective on algorithmic decisions. We predict that the perceived transparency of another human's decision processes would be lower than the perceived transparency of a consumer's own decision processes, but higher than perceived transparency for algorithmic decision processes, simply because humans can project their own decision processes onto other humans even though they do so with error (Dunning et al. 1990). Consequently, trust in and satisfaction with other's choice would be lower than with self choice, but higher than trust in and satisfaction with algorithmic choice. This expanded focus further highlights the role of perceived process transparency in accounting for algorithm acceptability.

In addition to our study of other human's decisions relative to self and algorithmic decisions, study 2 is conducted in a moral decision domain: a car accident. Moral considerations of algorithmic decisions are important, since algorithms may make intricate trade-offs with important ethical implications. In these cases, unlike in the decisions examined in the first two studies, the impact of the decision often involves other people beyond just the decision-maker. For instance, self-driving vehicles need to solve the dilemma of whether passengers or pedestrians should be sacrificed in a car accident (Awad et al. 2018; Bonnefon, Shariff, and Rahwan 2016). In 2019, 36,096 people died in car crashes in the US alone (U.S. Department of Transportation's National Highway Traffic Safety Administration 2020), or approximately 100 deaths per day. Most of these accidents never make headlines. Nonetheless, the death of a pedestrian involving Uber's self-driving car was covered in all major news sources (e.g., Bensinger and Higgins 2018; McFarland 2018; Wakabayashi 2018). One reason for why we are more disturbed by self-driving car accidents could be that we perceive the decision processes of self-driving cars as less transparent than our own or other human drivers' decisions. The nontransparency of the decision processes of self-driving cars has been discussed as a potential impediment to the trust in them (Shariff, Bonnefon, and Rahwan 2017).

Bridging the two contributions of study 2, we hypothesized that a self-driving algorithms' decisions would be considered less process transparent, trustworthy, and acceptable than another human driver's decisions, which would be viewed less favorably than decisions one makes when driving themselves (H1a). We expected that the mechanism for these effects is that the lower perceived process transparency of an algorithm (vs. a human) would lead to lower trust and, therefore, lower decision acceptability (H1b).

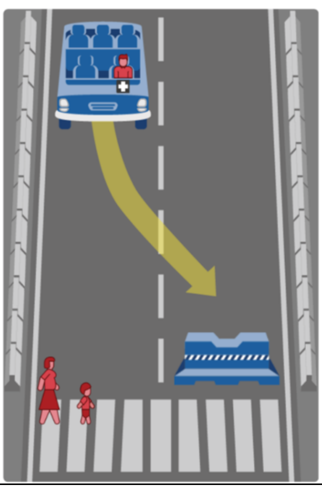
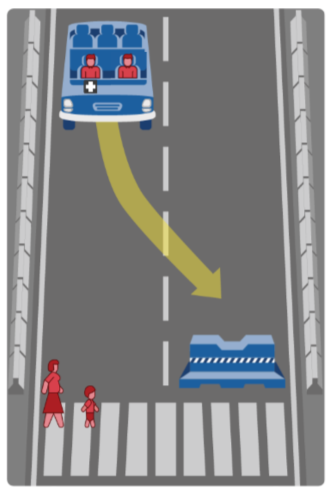
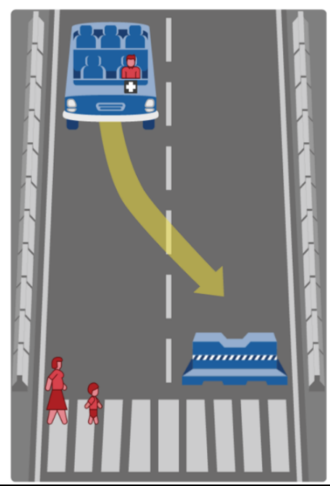
Method

Participants. We recruited participants who have a driver's license, regularly drive, and have not been in a car accident that has caused a serious injury to either themselves or another person. This was to ensure that we recruited individuals who are familiar with the context of this study, which was driving, and also individuals who would not be highly sensitive to the scenario, which was a car accident resulting in a severe injury. Participants were four hundred fifty-six MTurk workers that matched these criteria (M_{age} = 40.2, SD = 12.9, 58.8% female). They were randomly assigned to one of three conditions: Self Choice, Other's Choice, Algorithmic Choice.

Materials and Procedure. All three conditions involved an inevitable car accident. In order to design these scenarios, we used MIT's Moral Machine, an online platform developed to investigate how individuals would prefer self-driving cars to resolve life-or-death choices in an inevitable accident (Awad et al. 2017). Participants viewed a scenario where the car could crash into pedestrians or a concrete barrier and read a description of their assigned scenario (see Table

1 for these scenarios and corresponding visuals). Participants received minimal information on why the driver chose this path so that we could measure perceived process transparency.

TABLE 1 (CHAPTER 1). STUDY 2: EXPERIMENTAL CONDITIONS

Self Choice	Other's Choice	Algorithmic Choice
Imagine that you are driving your car to your friend's dinner party. Two pedestrians step into the crosswalk. The car swerves and crashes into a concrete barrier on the other side of the road, severely injuring you. 	Imagine that you are in an Uber, going to your friend's dinner party. Two pedestrians step into the crosswalk. The car swerves and crashes into a concrete barrier on the other side of the road, severely injuring you. 	Imagine that you are in a self-driving car, going to your friend's dinner party. Two pedestrians step into the crosswalk. The car swerves and crashes into a concrete barrier on the other lane, severely injuring you. 

After viewing their assigned scenario, the participants reported decision acceptability via 3 sub-items ("This decision was acceptable", "This decision meets my approval", "I welcome this decision"; Cronbach's alpha = .96) on a 7-point scale (1 = "Strongly disagree", 7 = "Strongly agree"). Next, they specified how much they trust either their own (SC), the driver's (OC), or the self-driving car algorithm's (AC) driving ability in this scenario on a 7-point scale

(1 = “Strongly distrust”, 7 = “Strongly trust”)[2]. Additionally, we measured their willingness to ride (WTR) in a car by asking how likely they would be to ride in a car driven by themselves (SC), another person (OC), or a self-driving car (AC), in the future, regardless of any concerns about accessing a vehicle (1 = “Very unlikely”, 7 = “Very likely”). Finally, as the process transparency measure for the decision made prior to the crash, they rated how well they understand the processes behind their (SC), the driver’s (OC), the self-driving car algorithm’s (AC) decision in this scenario on a 7-point scale (1 = “Little to no understanding”, 4 = “Moderate understanding”, 7 = “Detailed and deep understanding”).

Results

We conducted planned comparisons for the effect of choice type on our dependent variables (Table 2). We found that Self Choice was rated more favorably than Other’s Choice, which was rated more favorably than Algorithmic Choice for all our dependent variables: trust, willingness to ride, and perceived process transparency.

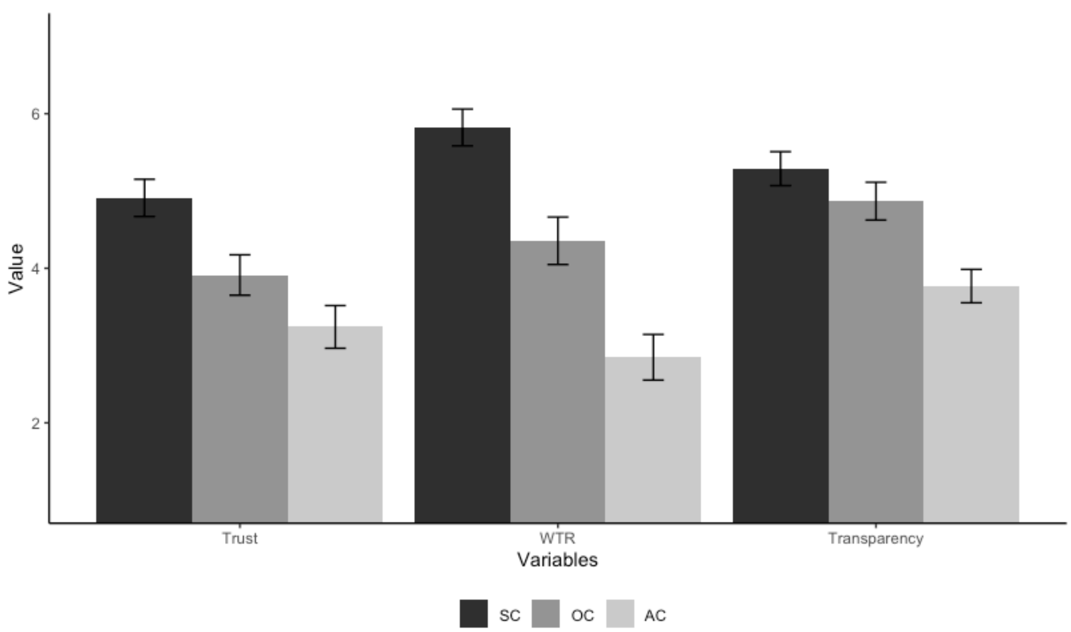
[2] Since decision acceptability and trust had a Cronbach’s α of .90, we combined the two measures under one trust index. Accordingly, the trust results we report are for the combination of these two measures.

TABLE 2 (CHAPTER 1). STUDY 2: CHOICE TYPE PLANNED COMPARISON FINDINGS

Planned Comparisons			F	p	R^2
Self Choice vs. Other's Choice	M_{SC}	M_{OC}			
Trust	4.91	3.91	25.52	< .001	.05
WTR	5.82	4.36	38.42	< .001	.08
Process Transparency	5.29	4.87	5.58	.019	.01
Self Choice vs. Algorithmic Choice	M_{SC}	M_{AC}			
Trust	4.91	3.24	79.33	< .001	.15
WTR	5.82	2.85	214.20	< .001	.32
Process Transparency	5.29	3.77	85.21	<. 001	.16
Other's Choice vs. Algorithmic Choice	M_{OC}	M_{AC}			
Trust	3.91	3.24	11.17	< .001	.02
WTR	4.36	2.85	40.70	< .001	.08
Process Transparency	4.87	3.77	40.87	< .001	.08

Note. Degrees of freedom in these analyses are 1, 454.

FIGURE 5 (CHAPTER 1). STUDY 2 RESULTS



Note. Error bars represent ± SEM.

In order to test the process of how choice type, specifically Other's Choice versus Algorithmic Choice, affects trust via perceived process transparency, we conducted a mediation analysis using a bootstrap procedure with 10,000 samples (Model 4 of Hayes 2013). We found an indirect-only mediation with an indirect effect of .344, with standard error .057, 95% CI = [.235, .461]. Algorithmic (vs. other) choice was less process transparent (a = .549, SE = .086, p < .001) and hence less trustworthy (b = .626, SE = .041, p < .001). There was no direct effect (c = -.009, SE =.088, p = .923). This mediation was replicated with algorithmic (vs. self) choice as the independent variable, as in study 1b, and with self (vs. other's) choice as the independent variable (for these additional mediation analyses see Appendix A, Section 3).

Discussion

Algorithms that solve ethical dilemmas and act as moral agents in consequential decisions will become more prominent in the future, deciding important issues such as who lives or dies in a self-driving car accident. In study 2, we tested the influence of choice type and the role of process transparency in a car accident scenario, where algorithms may have a large influence on human lives.

We found that when there is a decision between crashing into pedestrians or a concrete barrier, people are more comfortable making this decision themselves than letting an algorithm or another human make this decision because they view their decision processes as more transparent and therefore trustworthy. In particular, our addition of another human's choice condition revealed that another human's decision was seen as more process transparent and trustworthy than algorithmic choice. Furthermore, we provided insight to the proposed mechanism, whereby algorithmic (vs. another human's) choice results in less perceived process transparency, which generates lower trust. Thus, study 2 provided support for our hypothesis that any human's (self or other) decision is viewed more favorably than algorithmic choice due to higher levels of perceived process transparency.

In the first three studies, we have measured perceived process transparency and found a strong relationship between process transparency and trust for human versus algorithmic choice in contexts where algorithmic decisions are mostly unfamiliar to consumers. In study 3, we aimed to see if the role of process transparency for explaining trust was similarly important in a situation where algorithmic decisions are comparatively familiar.

STUDY 3: NATURAL VARIATION IN TRUST AND ITS RELATION TO PROCESS TRANSPARENCY

Thus far, studies 1a, 1b, and 2 provided evidence for process transparency in explaining algorithm acceptance. In all of these scenarios, the algorithmic decision makers being investigated (date choice, gamble choice, and driving choice) could be considered relatively unfamiliar to most consumers. What about instances where algorithmic decision makers have already become standard, and are thus more familiar to consumers? Is the role of perceived process transparency for explaining trust equally important in a context where algorithm acceptance is relatively high? With this question in mind, we explored the role of perceived process transparency for algorithms that are liked versus disliked. Our prediction is that there will be a larger gap between self and algorithmic process transparency when consumers dislike (vs. like) algorithms. This natural variation in trust for known algorithms provides additional insight into the role of perceived process transparency as a component of that trust.

We conducted a pretest ($N = 250$, $M_{age} = 39.6$, SD = 11.1, 48.8% female) in which participants read descriptions of fourteen different domains where algorithms are available to replace human decisions (e.g., Google, Google Maps, Netflix, Spotify, Uber; see Appendix A, Section 3 for exact descriptions and additional details). After reading a domain's description, they were asked whether they preferred choosing themselves or letting the algorithm choose for them in that particular domain, as well as whether they trusted their or the algorithm's ability to choose for them more in that domain. Participants preferred and trusted self choice more than algorithmic choice in most of the domains we tested, with the exception of Google Maps. Accordingly, in study 3 we use the contexts of Google (a context where self choice was preferred and trusted more) and Google Maps (a context where algorithmic choice was preferred and trusted more) in order to study whether process transparency continues to explain trust in a context where algorithm aversion is observed, e.g., Google, than in a context where algorithm appreciation is observed, e.g., Google Maps.

Method

Participants. We recruited five hundred and eighty-eight Amazon Mechanical Turk (MTurk) workers ($M_{age} = 39.9$, SD = 12.0, 46.3% female). Participants were randomly assigned to one of four conditions in a 2 (Choice type: Self Choice (SC), Algorithmic Choice (AC)) x 2 (Choice context: Google, Google Maps) between-subjects design.

Materials and Procedure. All participants were asked to imagine that they need to pick up some nails from a local hardware store. Participants in the Google conditions were told that they went onto Google to search for a local hardware store. They saw a screenshot of the Google search bar with "local hardware store" typed in. Google SC participants were told that after clicking on Google Search, they see 3 store options that Google found and that they picked one of these options. In contrast, Google AC participants were informed that they clicked on "I'm Feeling Lucky", whereafter Google's algorithm chose a hardware store among search results.

Participants in the Google Maps conditions were told that they went onto Google Maps to search for directions to a local hardware store. They saw a screenshot of the Google Maps search bar with "local hardware store" typed in. Google Maps SC participants saw 3 route options that Google Maps found, and they were told that they picked one of the options. Google Maps AC participants saw the same 3 route options and also that the Google Maps algorithm picked one of the route options for them (see Appendix A, Section 4 for further details on the stimuli).

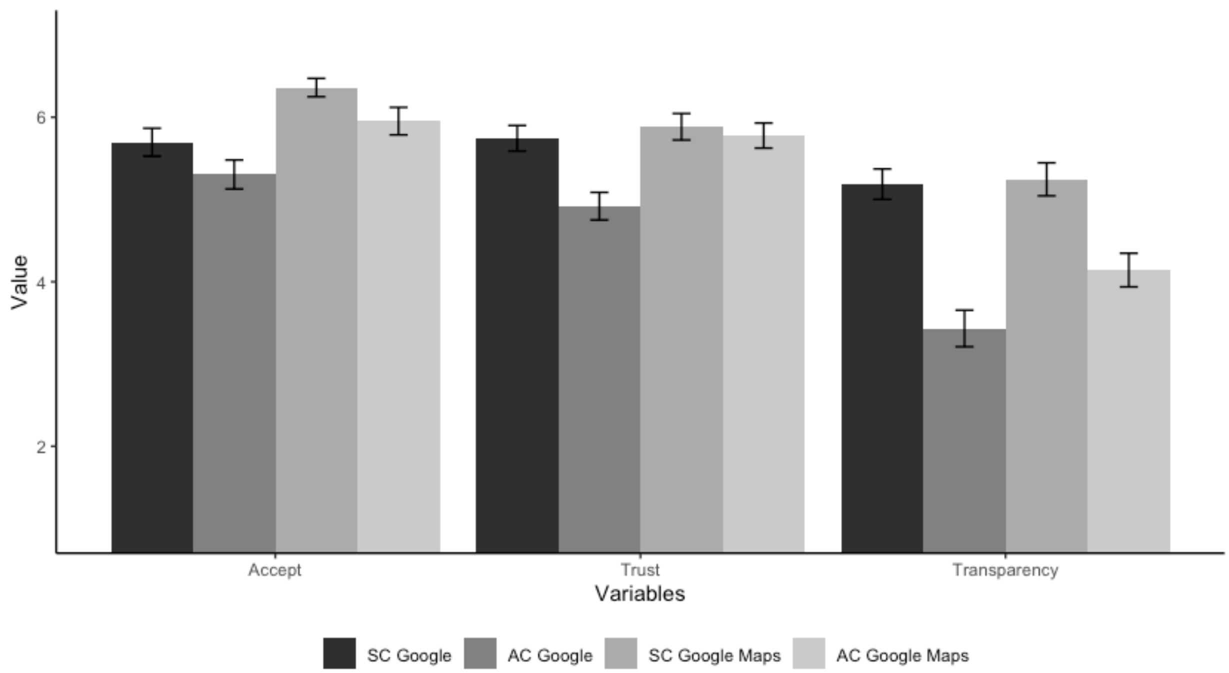
Participants were then asked to fill out the outcome measures. First, they rated the decision acceptance measure by indicating how willing they would be to go to the chosen store (Google) or go to the store by using the chosen route (Google Maps) on a 7-point scale (1 = "Very unwilling", 7 = "Very willing"). Second, they reported how much they trusted their (SC) or the algorithm's (AC) ability in deciding on which store to go to (Google) or which route to take to the store (Google Maps) on a 7-point scale (1 = "Strongly distrust", 7 = "Strongly trust"). Then, as a perceived accuracy measure[3], they rated how accurate they think they or the algorithm would be at choosing the best store (Google) or the best route (Google Maps) on a 7-point scale (1 = "Very inaccurate", 7 = "Very accurate"). Finally, as the process transparency measure, they

[3] We find in a confirmatory factor analysis that the perceived accuracy measure is not distinguishable from our existing trust measure, suggesting that they represent the same basic construct. We conclude that perceived accuracy is indeed a concern for consumers when considering algorithmic choice but it is a crucial element of overall trust and therefore we do not discuss it as a separate construct in our studies. More details are available in the appendix.

indicated how well they understand the processes behind their or the algorithm's choice on a 7-point scale (1 = "Little to no understanding", 4 = "Moderate understanding", 7 = "Detailed and deep understanding").

Results

FIGURE 6. STUDY 3 RESULTS



Note. Error bars represent ± SEM.

In order to test our hypothesis that the effect of choice type on our dependent variables is greater for Google than Google Maps, we separately regressed decision acceptance, trust, and process transparency on choice type (Self Choice = 1, Algorithmic Choice = -1), choice context (Google = 1, Google Maps = -1), and their interaction. First, a regression with decision acceptance was conducted ($F(3, 584) = 30.26$, $p < .001$, $R^2 = .14$). Decision acceptance was greater for SC than AC (B = .200, *SE* = .040, $t(584) = 4.95$, $p < .001$) and for Google Maps than Google (B = -.328, *SE* = .040, $t(584) = -8.12$, $p < .001$). The interaction was not significant.

Second, a regression with trust as the dependent variable was conducted ($F(3, 584) = 30.19$, $p < .001$, $R^2 = .13$). It revealed an effect of choice type (B = .233, *SE* = .041, $t(584) = 5.74$, $p < .001$) and choice context (B = -.249, *SE* = .041, $t(584) = -6.14$, $p < .001$), whereby algorithmic (vs. self) choice and Google (vs. Google Maps) each elicited lower trust. There was a choice type by choice context interaction (B = .180, *SE* = .041, $t(584) = 4.42$, $p < .001$). Further investigation of this interaction through contrasts revealed that SC and AC did not differ on trust for Google Maps ($F(1, 586) = .74$, $p = .391$, $R^2 = .001$), whereas trust was lower for AC than SC for Google ($F(1, 586) = 49.02$, $p < .001$, $R^2 = .08$).

The results for the regression with process transparency as the dependent variable ($F(3, 584) = 70.83$, $p < .001$, $R^2 = .26$) uncovered a significant effect of choice type (B = .714, *SE* = .052, $t(586) = 13.74$, $p < .001$) and choice context (B = -.192, *SE* = .052, $t(586) = -3.70$, $p < .001$), where algorithmic (vs. self) choice and Google (vs. Google Maps) decisions had lower process transparency. There was a choice type by choice context interaction (B = .163, *SE* = .052, $t(586) = 3.13$, $p = .002$), such that the effect of choice type on transparency was higher for Google (vs. Google Maps). Contrasts revealed that AC (vs. SC) was much less transparent for Google ($F(1, 586) = 127.70$, $p < .001$, $R^2 = .18$) than for Google Maps ($F(1, 586) = 44.62$, $p < .001$, $R^2 = .07$).

We conducted a mediation analysis using 10,000 bootstrapped samples in which we found that SC (vs. AC) led to higher process transparency ($a = .714$, SE = .053, $p < .001$), and therefore, higher trust ($b = .270$, SE = .030, $p < .001$). There was a significant indirect-only mediation of perceived process transparency for the effect of choice type on decision trust, indirect effect = .193, SE = .025, 95% CI = [.146, .244] (Model 4 of Hayes 2013).

A moderated mediation analysis using 10,000 bootstrapped samples revealed that the indirect effect of choice type on trust through process transparency was more substantial for Google than for Google Maps: estimated difference (ACME of Google - ACME of Google Maps) was .101, 95% CI: [.001, .203], $p = .047$ (Model 8 of Hayes 2013). In other words, process transparency mediated the relationship between choice type and trust more for Google than Google Maps.

Discussion

The results of study 3 offer additional support for H1a and H1b that perceived process transparency drives algorithmic trust and acceptance. Furthermore, we show that process transparency matters more for algorithmic trust when algorithms are disliked more and are less standard as decision makers (e.g., Google). When algorithmic decisions are equally trustworthy as consumers' own decisions (e.g., Google Maps), the role of process transparency in determining trust shrinks. While the relationships tested in this study may be seen as purely correlational, the natural variations in trust and liking for these different algorithms provide an opportunity to see how greater gaps between self and algorithmic process transparency relate to trust, satisfaction, and liking of algorithms. Overall, the first four studies indicate that perceived process transparency is a novel and important component of algorithm acceptance. This suggests that enhancing process transparency through an intervention can enhance algorithm acceptance. On that account, study 4 was designed to test such an intervention, namely, input explainability.

STUDY 4: INPUT EXPLAINABILITY TO ENHANCE PROCESS TRANSPARENCY (LOAN APPLICATION)

Studies 1a, 1b, and 2 found evidence of the perceived process transparency mechanism using existing levels of perceived process transparency within the study scenarios. Study 3 demonstrated the importance of process transparency for algorithms that are already liked and trusted versus when they are not. Study 4 was conducted in order to test whether an input explainability intervention, defined as enhancing consumers' ability to know relevant input information, can increase perceived process transparency in the context of a loan application scenario. We manipulated how explainable the decision was by making the inputs of the decision knowable or unknowable. We hypothesized that increased input explainability would increase perceived process transparency and trust, and hence satisfaction (H2a). We also hypothesized that this effect would be larger for algorithms than humans (H2b).

Method

Participants. We recruited nine hundred and one Amazon Mechanical Turk (MTurk) workers (M_{age} = 41.56, SD = 12.83, 51.2% female). The experiment followed a 2 (Choice Type: Other's Choice (OC), Algorithmic Choice (AC)) x 3 (Input Explainability: Explainable (X), Unexplainable (UnX), Baseline) between-subjects design.

Materials and Procedure. Participants read that they applied their bank for a loan and that their loan application was denied. Input explainable participants were provided with the inputs that went into bank agent's (OC) or bank algorithm's (AC) decision. Input unexplainable participants read that the agent (OC) or the algorithm (AC) based their decision on inputs that were strictly confidential. Baseline participants received no information about how the agent (OC) or the algorithm (AC) made this decision (see Table 3 for the stimuli).

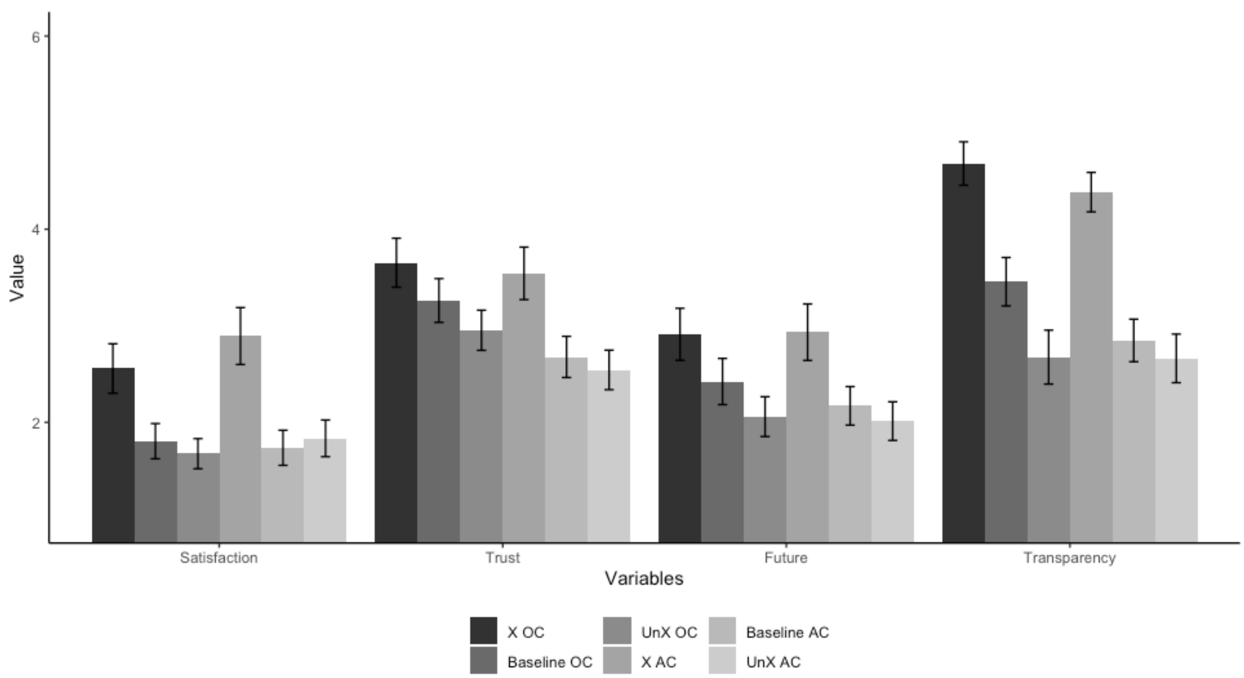
TABLE 3 (CHAPTER 1). STUDY 4: EXPERIMENTAL CONDITIONS

	Explainable	**Unexplainable**	**Baseline**
Other's Choice	You applied to your bank for a loan and your loan application was denied. The bank agent evaluated your loan application. The agent based their assessment on: 1. Your stability (how long you have been living at your current address and how long you have been in your current job) 2. Your debt-to-income ratio 3. The value of your assets 4. Your record of paying your bills on time and in their entirety	You applied to your bank for a loan and your loan application was denied. The bank agent evaluated your loan application. The agent based their assessment on inputs that are strictly confidential.	You applied to your bank for a loan and your loan application was denied. The bank agent evaluated your loan application.
Algorithmic Choice	You applied to your bank for a loan and your loan application was denied. The bank's algorithm evaluated your loan application. The algorithm based its assessment on: 1. Your stability (how long you have been living at your current address and how long you have been in your current job) 2. Your debt-to-income ratio 3. The value of your assets 4. Your record of paying your bills on time and in their entirety	You applied to your bank for a loan and your loan application was denied. The bank's algorithm evaluated your loan application. The algorithm based its assessment on inputs that are strictly confidential.	You applied to your bank for a loan and your loan application was denied. The bank's algorithm evaluated your loan application.

After reading their assigned scenario, participants rated how satisfied they are with this decision (1 = "Very dissatisfied", 7 = "Very satisfied"), how much they trust the bank agent's (OC) or the algorithm's (AC) ability to decide on this loan application (1 = "Strongly distrust", 7 = "Strongly trust"), how likely they are to use this bank's services in the future (1 = "Very unlikely”, 7 = “Very likely”). Finally, they indicated how well they understood the processes behind the bank agent’s (OC) or the bank algorithm’s (AC) decision to decline their loan application, which provided the process transparency measure (1 = “Little to no understanding”, 4 = “Moderate understanding”, 7 = “Detailed and deep understanding”).

Results

FIGURE 7. STUDY 4 RESULTS



Note. Error bars represent ± SEM.

The overall role of input explainability. Overall, we found that input explainable decisions were rated more favorable than baseline decisions, which were rated more favorable than input unexplainable decisions for all our dependent variables: satisfaction, trust, future use, and process transparency (Table 4).

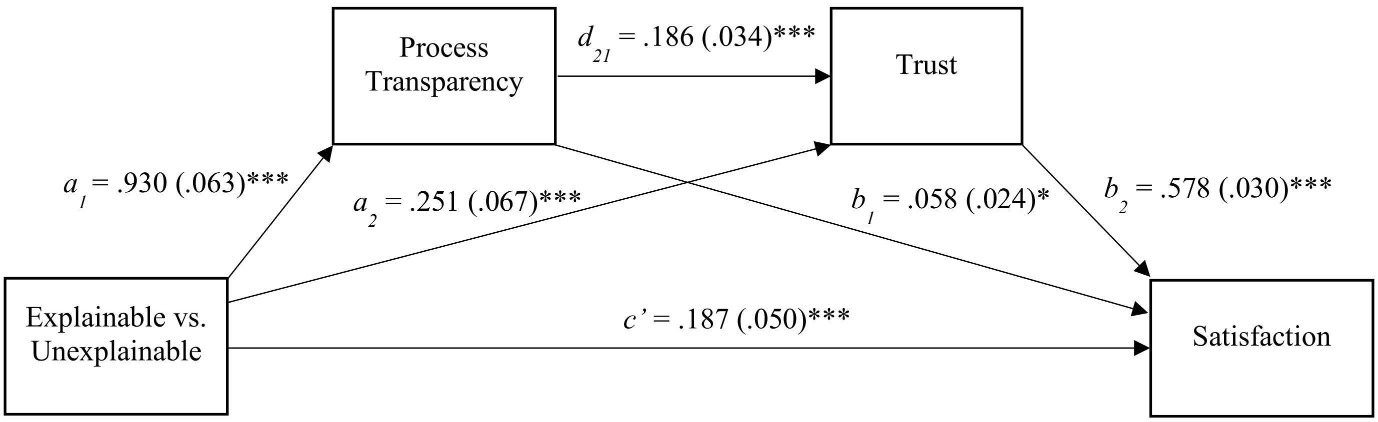
TABLE 4 (CHAPTER 1). STUDY 4: INPUT EXPLAINABILITY PLANNED COMPARISON FINDINGS

Planned Comparisons			F	p	R^2
Explainable vs. Unexplainable	M_X	M_{UnX}			
Satisfaction	2.73	1.75	75.77	< .001	.08
Trust	3.60	2.75	50.76	< .001	.05
Future	2.92	2.04	53.59	< .001	.06
Process Transparency	4.53	2.67	223.50	< .001	.20
Explainable vs. Baseline	M_X	M_{Base}			
Satisfaction	2.73	1.77	73.06	< .001	.08
Trust	3.60	2.97	27.16	< .001	.03
Future	2.92	2.30	25.89	< .001	.03
Process Transparency	4.53	3.15	110.00	<. 001	.11
Unexplainable vs. Baseline	M_{UnX}	M_{Base}			
Satisfaction	1.75	1.77	9.30	.884	< .001
Trust	2.75	2.97	10.93	.072	.004
Future	2.04	2.30	40.70	.036	.005
Process Transparency	2.67	3.15	40.87	< .001	.01

Note. Degrees of freedom in these analyses are 1, 899.

We tested our hypothesis of input explainability affecting satisfaction via perceived process transparency and trust in a serial mediation analysis using 10,000 bootstrapped samples (Figure 8). This analysis revealed an indirect effect ($a_1*d_{21}*b_2$) of .100 (SE = .021, 95% CI = [.064, .144]). We found support for our hypothesis that input explainable (vs. unexplainable) decisions are perceived as more process transparent and, hence, more trustworthy, leading to higher satisfaction, consistent with H2a. This finding was replicated in separate serial mediations with input explainable (vs. baseline) and input unexplainable (vs. baseline) as the independent variables, which correspondingly demonstrated that compared to the baseline, satisfaction can be enhanced by boosting explainability and be hurt by undermining explainability through perceived process transparency and trust (see Appendix A, Section 5 for details).

FIGURE 8 (CHAPTER 1). STUDY 4: INPUT EXPLAINABLE (VS. UNEXPLAINABLE) DECISIONS ARE MORE PROCESS TRANSPARENT AND, HENCE, MORE TRUSTWORTHY, LEADING TO HIGHER SATISFACTION



Note. * $p < .05$, *** $p < .001$. Parameter estimates and standard errors (in parentheses) are indicated. Indirect effect (a_1*d_{21}*b_2) = .100***, SE = .021, 95% CI = [.064, .144].

The differential effect of input explainability for humans and algorithms. In order to test H2b that the effect of boosting input explainability is greater for algorithmic decisions than human decisions, we separately regressed each of our dependent variables on choice type (Other's Choice = 1, Algorithmic Choice = -1), input explainability (Explainable = 1, Baseline = -1), and their interaction.

The results for the regression with satisfaction as the dependent variable ($F(3, 897) = 26.40$, $p < .001$, $R^2 = .08$) revealed a non-significant main effect of choice type (B = -.072, SE = .045, $t(897) = -1.59$, $p = .112$) and a significant main effect of input explainability (B = .477, SE = .056, $t(897) = 8.54$, $p < .001$), where input explainable (vs. baseline) decisions had higher satisfaction. There was a marginally significant choice type by explainability interaction (B = -.102, SE = .056, $t(897) = -1.82$, $p = .069$), such that the effect of input explainability on satisfaction was higher for algorithmic decisions ($F(1, 899) = 52.92$, $p < .001$, $R^2 = .06$) than for other's choice ($F(1, 899) = 21.11$, $p < .001$, $R^2 = .02$).

The regression with trust as the dependent variable ($F(3, 897) = 15.24$, $p < .001$, $R^2 = .05$) revealed a main effect of choice type (B = .183, $SE = .049$, $t(897) = 3.75$, $p < .001$) and input explainability (B = .313, $SE = .060$, $t(897) = 5.24$, $p < .001$), whereby other's (vs. algorithmic) choice and explainable (vs. baseline) decisions each elicited higher trust. Consistent with our hypothesis, there was a choice type by explainability interaction (B = -.119, $SE = .060$, $t(897) = -1.99$, $p = .047$), in which increased input explainability had a greater impact on enhancing trust in algorithmic choice ($F(1, 899) = 25.99$, $p < .001$, $R^2 = .03$) than other's choice ($F(1, 899) = 4.95$, $p = .026$, $R^2 = .005$).

While we did not have an explicit hypothesis, in addition to investigating the differential benefit of boosting input explainability for algorithmic (other's) choice (per H2b), we wanted to study the differential harm reducing input explainability may have on human choice. Accordingly, we regressed our dependent variables on choice type (Other Choice = 1, Algorithmic Choice = -1), unexplainability (Unexplainable = 1, Baseline = -1), and their interaction ($F(3, 897) = 8.14$, $p < .001$, $R^2 = .03$). Participants perceived process transparency of other's (vs. algorithmic) choice higher (B = .150, $SE = .056$, $t(897) = 2.68$, $p = .008$) and unexplainable (vs. baseline) decisions as lower (B = -.243, $SE = .068$, $t(897) = -3.55$, $p < .001$). The interaction effect (B = -.151, $SE = .068$, $t(897) = -2.21$, $p = .027$) indicated that reduced input unexplainability harmed process transparency for other's choice ($F(1, 899) = 16.28$, $p < .001$, $R^2 = .02$), whereas it did not affect algorithmic choice ($F(1, 899) = .866$, $p = .352$, $R^2 < .001$). There were no other interactions between input explainability and choice type.

Discussion

Study 4 establishes input explainability as an effective intervention to increase perceived process transparency, as well as to increase trust and satisfaction, for all types of decision makers. Moreover, we confirmed the proposed relationship in H2a: input explainability enhances perceived process transparency, and therefore, trust, which augments satisfaction. This suggests that consumers inferred the underlying decision process through the provision of input information. That is, the knowledge about *why* a decision was made led to inferences about *how* a decision was made. This carries important implications since it suggests that firms implementing algorithmic support for consumers and wanting to amplify trust and satisfaction can simply share information on inputs to both human and algorithmic decision makers rather than attempt to explain their internal workings.

The effects of input explainability are especially notable for algorithms, supporting H2b. Based on our findings from the previous studies where algorithms were perceived as less process transparent than humans at baseline, we investigated the differential effects of augmenting input explainability and undermining input explainability compared to the baseline for both human choice and algorithmic choice. We found that input explainability influenced human choice and algorithmic choice differently. Specifically, boosting input explainability enhanced trust and satisfaction more for the algorithmic decisions than the human decisions, since the human decisions were already perceived as more process transparent by default. In addition, undermining input explainability did not change the process transparency of algorithms, which were already low on process transparency by default, but did hurt the perceived process transparency of other humans. Hence, increasing input explainability relative to baseline information is a more effective intervention for algorithmic decisions than human decisions.

GENERAL DISCUSSION

Theoretical Contributions

In five studies across a variety of consumer decision domains, we test possible drivers for the general distrust in, and occasional acceptance of, algorithmic decisions. We find that algorithm aversion is affected by a perception that algorithms are less process transparent than our own decision-making processes (studies 1a, 1b, and 2). Lower perceived process transparency for algorithms leads to less trust and less satisfaction than human decisions for the same choice outcome. It is worth noting that we also find that algorithmic decisions are consistently judged to be less able to account for the decision maker's unique tastes, yet this uniqueness account does not mediate trust or satisfaction in our contexts, probably because the outcomes themselves do not reflect unique preferences. While not directly tested, the persistence of algorithm aversion due to lower perceived process transparency even in situations where outcomes are identical and little process information is given, such as in Study 2, adds to the literature showing that consumers believe they have insight into their own decision processes even when they do not (Dunning 2011; Nisbett and Wilson 1977; Wilson 2004).

Our findings also provide marketers who use algorithms a way to make those algorithms more trustworthy through input explainability. Providing consumers with information about inputs makes decision makers, both human and algorithmic, seem more process transparent and thus more trustworthy. This offers an important option for increasing process transparency without actually explaining the inner workings of an algorithm. In a sense, consumers are operating under the illusion of an explanation of *how* (process) when all they have received is an explanation of *why* (inputs). This is consistent with explanation-giving in other contexts, in which individuals accept decisions with explanations even when the explanation offers no true information (Langer et al. 1978).

Throughout this paper, and in all of our studies, we have focused on perceived process transparency, trust, and satisfaction for algorithmic and human choices (self or other) – decisions made between multiple options. While our focus is on decisions, partly through a recognition that decisions made by algorithms on a consumer's behalf will continue to expand into additional aspects of daily life, we also acknowledge that the same conceptual model applies to forecasts and predictions made by algorithms for consumers. Thus, we predict that our findings about increasing input explainability as a way to increase perceived process transparency for algorithms may apply equally well to forecasting and recommendation environments.

Practical Implications

The current research suggests multiple fruitful courses of action for marketers. First, marketers should implement input explainability to enhance perceived process transparency of the algorithmic recommendations and decisions they are making on their digital platforms. For example, when customers search for an item they would like to purchase on Amazon, search results will often include an option with the Amazon's Choice badge. This algorithm-based choice is likely determined by something other than (or in addition to) popularity or ratings, since the badged option is rarely the bestseller or the highest rated alternative. The only entity that is privy to the facts behind this badge assignment is Amazon, which declines to describe the process behind Amazon's Choice (Matsakis 2019). Thus, Amazon's Choice is selected through a nontransparent process from the consumers' perspective. Our research suggests that consumers do not appreciate nontransparency since it reduces their trust in algorithmic decisions or recommendations. Digital marketplace companies like Amazon could implement input explainability as a method to augment process transparency, which would result in increased trust and, consequently, satisfaction and continued usage of their platforms. Note that doing so would mean that Amazon share some information into the inputs for Amazon's Choice, but not necessarily share how exactly an item is selected for the badge.

Second, as lawmakers have recognized that consumers have a desire to understand how an algorithm reaches a certain decision, they have developed regulations to address this. The Equal Credit Opportunity Act (ECOA) in the Unites States obliges creditors to inform applicants of "the specific reasons for the adverse action taken" (The United States Congress 2011). More recently, a prominent section of the European Union's General Data Protection Regulation (GDPR), Recital 71, emphasizes that a data subject has a right to "obtain an explanation of the decision reached" by automated processing (European Parliament and Council of the European Union 2016). GDPR's Article 13 also mandates data subjects receive "meaningful information about the logic involved" in automated decision-making. These legislations might be considered as explainability manipulations. By means of these legislations, consumers can request explanations for the decisions that algorithms make and possibly attain algorithmic process transparency and trust. Nonetheless, they only apply to when consumers receive decisions that have an unfavorable outcome in the ECOA's case and legal or similarly significant effects in the

GDPR's case, such as an online credit card application automatically being declined. Our findings from study 4 indicate that these interventions will be successful in enhancing algorithmic process transparency, trust, and satisfaction if they disclose relevant input information even for negative decisions like the loan application in our study. Therefore, future legislation or amendments to existing legislation would benefit from broadening their scope to consumers' right to input explanation for any algorithmic decision. Future legislation could even go a step further by providing input explainability as a built-in feature of algorithmic decisions such that an explanation is provided to consumers even before they request one. This would eliminate the inconvenience of going through the bureaucracy of requesting an explanation of the decision reached, resulting in a more trusted and satisfactory consumer-algorithm interactions.

Future Research

Since the current article offers a novel investigation into an incipient research area in consumer research, it motivates many exciting research questions. While we demonstrate that providing simple input information about a decision enhances consumers' trust and satisfaction in that decision, it would be valuable to continue to investigate whether and how the effect of input explainability can be moderated. In follow-up studies, we initiated some inquiries that would be fruitful to be pursued further in future research.

First, consumer expertise might moderate input explainability. Past research has shown that experts rely less on algorithmic advice than lay people (Logg et al. 2019). Experts might find an algorithm's input explainability less valuable for increasing satisfaction with the product, as they may already know how such algorithms make decisions. In an ancillary study, we tested this prediction in an algorithmic song selection decision on a music streaming platform, Spotify (for details, please see Appendix A, Section 6). We found that Spotify usage moderated the influence of input explainability on satisfaction. Specifically, our input explainability intervention was effective in increasing satisfaction for the 56.6% of participants who used Spotify more than 3.92 hours but less than 19.62 hours. These findings demonstrate a surprising result: the mid-level users appear to benefit from input explainability the most. Nonexperts are (dis)satisfied with and experts are satisfied with the algorithmic choice similarly whether it is explainable or unexplainable: explainability had no effect on experts or nonexperts. Therefore, companies should focus their explainability interventions on their mid-level users, as those customers may benefit more from such interventions.

Second, procedural fairness, the judgment about the fairness of a decision process (Blader and Chen 2011), may moderate input explainability. In an ancillary study, we found that input explainability increased trust for procedurally fair decisions, but decreased trust for unfair decisions (for details, please see Appendix A, Section 7). Input explainability also decreased satisfaction for procedurally unfair, but not fair, decisions. Therefore, we have initial evidence indicating that procedural fairness moderates the positive effect that input explainability has on trust and satisfaction such that if the procedures are seen as unfair, then additional knowledge of the inputs hurts trust and satisfaction. In contrast, if the process is demonstrated to be fair, then input explainability further boosts trust and satisfaction, even when outcomes are negative as in our Study 4. Future research can build on these preliminary findings and investigate the factors, such as procedural fairness, that companies need to implement into their decision-making systems in order for an input explainability intervention to be successful.

Future research can also address when and why knowing the inputs might substitute for understanding the decision process. Our work shows that input explainability can enhance consumers' ability to perceive the process transparency of a human or an algorithmic decision, so that the decision process is interpretable to them. This has important implications for the black box algorithms that have incomprehensible procedures even to their creators. Inputs, which are knowable, can augment perceived process transparency, even when that process is objectively opaque. Furthermore, enhanced perceived process transparency through input explainability would be preferable from companies' perspective as well, as algorithms are companies' intellectual property, whose source code and actual decision-making process, if knowable, are often kept secret. Therefore, our research demonstrates how consumers may lack an understanding of how algorithms operate but that they can infer their decision-making process through the provision of input information. Nonetheless, we do not yet know when and why knowing the inputs replaces the objective knowledge of the decision process. We believe investigating why exactly this occurs, for both algorithmic and other human decision makers, would prove to be valuable in future research.

Conclusion

Algorithms compel us to reimagine consumer decision-making. Since consumers have divergent conceptualizations of the process transparency of their own and algorithmic decision processes, they can be wary of algorithms making decisions on their behalf. The current research provides a method to boost satisfaction with the decisions, recommendations, and forecasts reached in the digital world: make those decisions input explainable and, therefore, process transparent and trustworthy. As algorithm-assisted and automated decision-making is becoming more widespread, this article suggests that digital platforms should prioritize providing input explanations for their black box procedures in order to promote trust in their algorithms and contentment with the goods and services they provide.

CHAPTER 2: BIAS TOLERANCE: WHEN HUMAN BIAS, BUT NOT ALGORITHMIC BIAS, IS DISREGARDED

ABSTRACT

People might regard humans as emotional and biased decision-makers, whereas they see algorithms as unemotional and neutral entities. Given that perceived bias may be expected to undermine trust in decisions, this paper addresses whether or not the perception of humans as biased affects trust in and satisfaction with human relative to algorithmic decisions. Across 4 studies (combined N = 3,121), we reveal a "bias tolerance" phenomenon, i.e., people acknowledge but disregard human bias and trust human decisions more than algorithmic ones. This bias tolerance effect in decision-making seems to occur because human emotionality strongly enhances trust and satisfaction for most decisions; the effect does not exist in data handling or in decisions for material purchases where emotionality is less valued.

Bias is a psychological mechanism which leads to systematic errors in judgements (Kahneman et al. 2021). It can manifest itself in societally important ways through systemic and unfair discrimination against specific individuals or groups of individuals in favor of others (Friedman and Nissenbaum 1996). For example, a bank may refuse to lend out a loan based on racial biases. Additionally, bias can indicate someone preferring a particular outcome due to reasons such as self-interested motives, familiarity, primacy or recency effects, and/or priming. For example, a consumer might choose to recommend a hotel they stayed at to another due to recency effects. All these types of psychological bias may lead to worse decisions, since they can cause systematic deviations from the fully rational considerations and the decision maker may reach suboptimal outcomes based on unreasonable grounds.

As decision makers, humans are biased (i.e., they have systematic errors of judgement) for the reasons described above. In addition, however, they are noisy. That is, there is unwanted variability of error among people who are expected to agree (Kahneman et al. 2021). In contrast to humans, algorithms as decision makers are not noisy since they will reproduce the same decision if the inputs are the same. Algorithms are thus a useful way of eliminating inconsistencies that human decision makers exhibit. Algorithms and even random linear models have been shown to make more accurate and consistent predictions than humans (Camerer 1981; Dawes 1979; Dawes and Corrigan 1974; Dawes et al. 1989; Grove et al. 2000; Meehl 1954; Yu and Kuncel 2020). Nonetheless, high accuracy and low noise does not mean that algorithms are unbiased.

Recent work has found that biased decision making can extend to algorithms as the training of the algorithms is often based on modeling the decisions and actions of biased people (e.g., bootstrapped models). Algorithms are often biased because the data that the algorithms use comes from a biased society (Barocas and Selbst 2016). Human bias is transmitted to or even augmented through algorithms by various means such as training data and data preparation. Therefore, algorithmic decisions can often reflect existing discrimination, especially if they are trained on past decisions. For instance, Amazon's hiring algorithm had to be terminated since its training data came from the hiring patterns in a male dominant technology sector (Dastin 2018); the algorithm preferred male applicants over female applicants. Thus, as algorithms substitute for human processes, they can still be biased. However, since they apply logical rules without emotionality across all decision scenarios, they are consistent, i.e., not noisy. Consumers view machines as cold and unemotional, while regarding humans as warm and emotional decision makers (Haslam et al. 2008). The lack of noise and emotionality in algorithmic decisions (Kahneman et al. 2021), might lead most individuals to see them as less biased than humans in many contexts.

Although observers may often consider humans as more biased than algorithms, this paper addresses whether or not this perception diminishes trust in and satisfaction with human (vs. algorithmic) decisions. Through a series of four studies utilizing a variety of decision contexts, we ask how people perceive levels of bias among humans versus among algorithms and whether these perceptions of bias affect trust and satisfaction with decisions. We reveal a "bias tolerance" phenomenon, i.e., consumers disregard human bias and trust human (vs. algorithmic) decisions.

Across four studies we uncover bias tolerance and explore boundaries for it. In study 1, we document bias tolerance and confirm its presence across both objective and subjective types of decisions. In study 2, we investigate and find that purchase type (i.e., experiential or material) moderates this relationship and that the effect is stronger for experiential purchases. In study 3, we find that consumers value emotionality more than bias and that this is why they tolerate bias among human decision makers. We test whether we can find environments where human bias is severe enough (e.g., unfair decisions with unfavorable outcomes) that human bias tolerance disappears, and algorithms become preferred. In contrast to increasing bias among humans, in study 4 we try to change the perceived bias of algorithms and study if there are tasks where the lack of bias among algorithms can make them more preferred. In situations that require computational skills and do not bring in emotionality, such as data handling, consumers might think that there is less chance for human bias to be involved. Thus, they may think that algorithms and humans are similarly biased, which might eliminate bias tolerance in tasks that do not necessitate human abilities, where emotionality may be valued.

STUDY 1: SUBJECTIVE AND OBJECTIVE DECISIONS

Study 1 (*N* = 640, Amazon Mechanical Turk) used a 2 (Decision type: Human, Algorithm) x 2 (Decision context: Loan application, Employee bonus) between-subjects design. The different scenarios of loan applications and employee decisions were employed since we aimed to test our hypotheses in contexts with varying degrees of objectivity. Consumers trust and use algorithms more for objective (e.g., financial) than subjective (e.g., dating) tasks (Castelo et al. 2019). Other work has demonstrated that algorithmic decisions are viewed as less fair than identical decisions by humans in the case of a subjective, employee bonus allocation context (Newman, Fast, and Harmon 2020). As we are interested in perceived bias, we wanted to conduct our study in both an objective context and a subjective context (financial vs. HR decisions) to address these findings and examine our research question in tasks that differ in terms of objectivity. Loan application participants read that they applied for a bank loan and that their application was evaluated either by financial advisors or an algorithm. Employee bonus participants read that their bonus was determined by either human resources or an algorithm (see Table 5 for details).

TABLE 5 (CHAPTER 2). STUDY 1: EXPERIMENTAL STIMULI

	Financial	**HR**
Human	Imagine that you seek to obtain a loan and you apply a bank. In order to determine your loan allocation, your bank relied on a financial advising team, which took into account a variety of factors. After the advising team made a series of deliberations, it determined the loan allocated to you.	Imagine the company you are working for just went through the process of making its end of the year bonus for each employee. In order to determine the size of the bonuses, your company relied on its human resources team, which took into account a variety of factors. After the human resources team made a series of deliberations, it determined the employee bonus allocated to you.
Algorithm	Imagine that you seek to obtain a loan and you apply a bank. In order to determine your loan allocation, your bank relied on an algorithm that took into account a variety of factors. After the algorithm made a series of computations, it determined the loan allocated to you.	Imagine the company you are working for just went through the process of making its end of the year bonus for each employee. In order to determine the size of the bonuses, your company relied on an algorithm that took into account a variety of factors. After the algorithm made a series of computations, it determined the employee bonus allocated to you.

Method

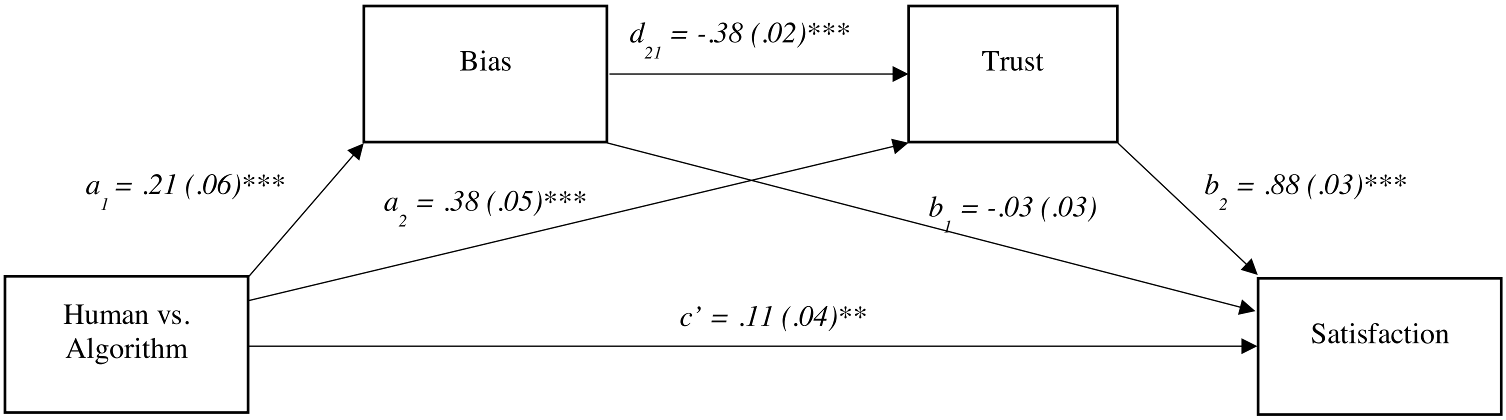
After reading their assigned scenario, participants indicated how satisfied they are with the bank's/company's decision procedure (1 = "Very dissatisfied", 7 = "Very satisfied"), how much they trust this bank's/company's ability to decide on this loan application (1 = "Strongly distrust", 7 = "Strongly trust"), and how biased they think this loan/bonus decision is (1 = "Very unbiased", 7 = "Very biased"). They also indicated how objective or subjective they think this decision was (1 = "Very subjective", 7 = "Very objective").

Results and Discussion

We predicted and found that the loan application (vs. bonus allocation) context was viewed as more objective: $F(1, 640) = 5.43$, $p = .020$. Across both scenarios, algorithmic decisions were deemed less satisfactory ($F(1, 640) = 37.98$, $p < .001$), trustworthy ($F(1, 640) = 29.35$, $p < .001$) and biased ($F(1, 640) = 12.28$, $p < .001$) than human decisions. There was no (Decision type) x (Decision context) interaction. This does not support Castelo et al.'s findings which was that consumers trust and use algorithms more for objective than subjective tasks.

A serial mediation analysis (Figure 9) uncovered an indirect effect of the human (vs. algorithm) decision on satisfaction through bias and trust, indirect effect = -.073, 95% CI = [-.119, -.032]: human decisions were perceived as more satisfactory through higher bias and trust. We revealed a bias tolerance effect in which perceiving humans as more biased than algorithms had no effect on satisfaction, whereas trust did. That is, despite high perceived bias, human decisions were more satisfactory because bias did not have a significant effect on satisfaction, while trust had a large positive effect on satisfaction. This was true in a more objective decision task (loan application) as well as a less objective decision task (bonus allocation).

FIGURE 9 (CHAPTER 2). STUDY 1: SERIAL MEDIATION ANALYSIS RESULTS



Note: * $p < .05$, *** $p < .001$. Parameter estimates and standard errors (in parentheses) are indicated. Indirect effect (a_1*d_{21}*b_2) = -.073***, 95% CI = [-.119, -.032]

Discussion

In study 1, we find that humans are viewed as more biased than algorithms. Nonetheless, this awareness does not lead to lower trust in and satisfaction with algorithmic decisions in tasks that varied by objectivity. Instead, the more biased human decisions are considered more trustworthy and more satisfactory. We define this relationship between perceived bias, trust, and satisfaction as an example of bias tolerance in human (vs. algorithmic) decisions. In study 2, we aimed to see if purchase type, that is if it is a material or an experiential purchase, moderates whether we observe bias tolerance.

STUDY 2: MATERIAL VS. EXPERIENTIAL PURCHASE MODERATOR

In study 2, we examined the differing effects of material versus experiential purchases on bias tolerance. Material purchases are “those made with the primary intention of acquiring a material good”; experiential purchases are “those made with the primary intention of acquiring a life experience” (Van Boven and Gilovich 2003, 1194). Material purchases are more comparable than experiential purchases (Carter and Gilovich 2010). Additionally, material goods are assessed based more on quality, whereas experiences are assessed based more on taste (Dai, Chan, and Mogilner 2020; Spiller and Belogolova 2017). Since experiences are less comparable and evaluated based on taste, human bias may play a bigger role in experiential (vs. material) purchases in explaining decision satisfaction. That is, given that we found in study 1 that high levels of bias among human decision makers can still lead to higher trust and satisfaction, we propose that consumers may think that human (vs. algorithmic) decisions are even more trustworthy and satisfactory due to higher bias when the purchases are experiential rather than material. In other words, we expect that bias tolerance might be attenuated in material (vs. experiential) purchases.

Method

We recruited 560 MTurk participants who were randomly assigned to one of four conditions in a 2 (Decision type: Human, Algorithm) x 2 (Purchase type: Material, Experiential) between-subjects design.

Depending on their assigned purchase type condition, participants were asked to state a material or an experiential purchase they are likely to make in the next year for which they will pay more than $50 (adapted from Dai, Chan, and Mogilner 2020). They were also provided with a definition of their assigned purchase type. Next, they read that either a person or an algorithm (depending on their assigned decision type) made a decision between all possible options on their behalf (see Table 6 for details).

Participants then completed the dependent variable measures. They reported the extent they would trust the person/algorithm's purchase decision (1 = "Strongly distrust", 7 = "Strongly trust"), the extent they would be satisfied with this person/algorithm's purchase decision (1 = "Strongly dissatisfied", 7 = "Strongly satisfied"), and how helpful they think this person/algorithm will be in making this purchase decision for them (1 = "Not at all", 7 = "A great extent"). Then they rated how biased they think this person/algorithm's purchase decision will be is (1 = "Very unbiased", 7 = "Very biased") and completed the manipulation check by indicating the extent to which the purchase they reported was material or experiential (1 = "Purely material," 9 = "Purely experiential").

TABLE 6 (CHAPTER 2). STUDY 2: EXPERIMENTAL STIMULI

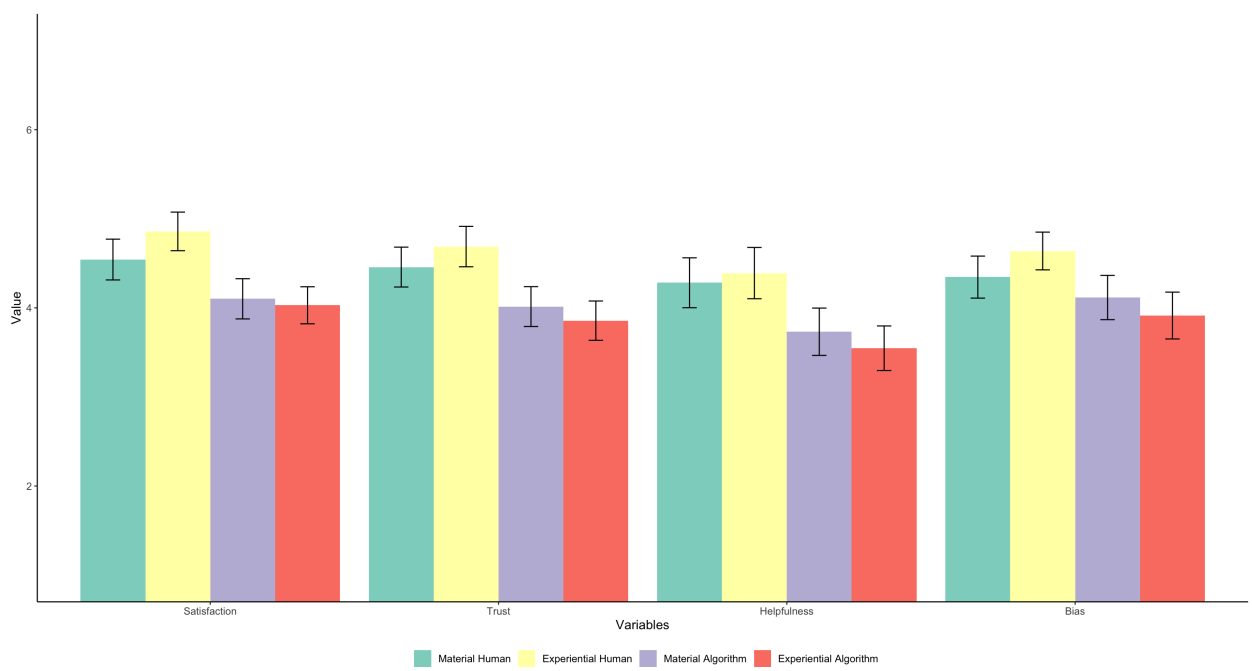
	Human	Algorithm
Experiential	Please think of an experiential purchase that you are very likely to make in the next year for which you will pay more than $50. By experiential purchase, we mean a purchase that involves spending money with the primary intention of acquiring a life experience – an event or series of events that you personally will encounter or live through. For example, vacation packages, meals at restaurants, and music and theater performances can be experiential purchases. Now imagine that a person who is knowledgeable about this product category made a decision between all possible options on your behalf.	Please think of an experiential purchase that you are very likely to make in the next year for which you will pay more than $50. By experiential purchase, we mean a purchase that involves spending money with the primary intention of acquiring a life experience – an event or series of events that you personally will encounter or live through. For example, vacation packages, meals at restaurants, and music and theater performances can be experiential purchases. Now imagine that an algorithm that is knowledgeable about this product category made a decision between all possible options on your behalf.
Material	Please think of a material purchase that you are very likely to make in the next year for which you will pay more than $50. By material purchase, we mean a purchase that involves spending money with the primary intention of acquiring a material possession – a tangible object that you obtain and keep in your possession. For example, clothing, jewelry, and various types of electronic gadgets can be material purchases. Now imagine that a person who is knowledgeable about this product category made a decision between all possible options on your behalf.	Please think of a material purchase that you are very likely to make in the next year for which you will pay more than $50. By material purchase, we mean a purchase that involves spending money with the primary intention of acquiring a material possession – a tangible object that you obtain and keep in your possession. For example, clothing, jewelry, and various types of electronic gadgets can be material purchases. Now imagine that an algorithm that is knowledgeable about this product category made a decision between all possible options on your behalf.

Results

Manipulation Check. Participants in the experiential (vs. material) condition rated their purchase as more experiential ($F(1, 558) = 1229$, $p < .001$). There was no other significant difference between experiential and material conditions.

Dependent Variables. Planned comparisons showed that, compared to algorithms, humans were seen as more satisfactory ($F(1, 558) = 31.84$, $p < .001$), trustworthy ($F(1, 558) = 31.11$, $p < .001$), helpful ($F(1, 558) = 25.33$, $p < .001$), and biased ($F(1, 558) = 15.06$, $p < .001$).

FIGURE 10 (CHAPTER 2). STUDY 2: RESULTS



Mediations. Consistent with previous studies, we find bias tolerance. Our serial mediation analysis where we investigated the influence of decision-maker type on satisfaction through bias and trust showed an indirect only mediation (indirect effect = -.03, SE = .012, 95% CI = [-.058, -

.011]) and bias had no direct impact on satisfaction: $b_1 = -.042$, SE = .022, 95% CI = [-.086, .001].

TABLE 7 (CHAPTER 2). STUDY 2: SERIAL MEDIATION ANALYSIS RESULTS

	Label	Estimation	SE
Human vs. Algorithm → Bias	a_1	.24***	.06
Human vs. Algorithm → Trust	a_2	.35***	.06
Bias → Trust	d_{21}	-.15***	.04
Human vs. Algorithm → Satisfaction	c'	.06	.03
Bias → Satisfaction	b_1	-.04	.02
Trust → Satisfaction	b_2	.85***	.02
Indirect Effect	$a_1 * d_{21} * b_2$	-.03*	.01

Note: * $p < .05$, ** $p < .01$, *** $p < .001$.

Additionally, a moderated mediation analysis (Hayes Model 7) revealed that the influence of decision-maker type (human versus algorithm) on $\text{Satisfaction}_{\text{index}}$[4] through bias was more important for experiential (vs. material) purchases: ACME (Material) – ACME (Experiential) = .039, 95% CI = [.0009, .0876]. Importantly, the conditional indirect effect of choice type on satisfaction index through bias was not significant for material purchases (effect = -.0181, 95% CI = [-.0532, .0090]), whereas it was significant for experiential purchases (effect = -.0574, 95% CI = [-.1026, -.0206]). In other words, when participants were planning experiential purchases, human (vs. algorithmic) decisions were viewed as more satisfactory due to higher bias. However, when participants were planning material purchases, higher satisfaction with a human (vs. algorithmic) decision-maker was not explained by higher bias.

[4] To make the moderated mediation simpler, combined trust, satisfaction, and helpfulness variables under one satisfaction index (Cronbach's α = .92).

Discussion

In study 2, we find that human bias explains the relationship between choice type and satisfaction in experiential purchases, but not in material purchases. That is to say, higher human (vs. algorithmic) bias was acknowledged in experiential purchases, however, that did not result in lower satisfaction as we might have expected. Study 3 was conducted to determine why humans are more trustworthy and satisfactory, despite being more biased, consequently demonstrating bias tolerance. We also asked could we design a situation where human bias is so strong that it diminishes trust? Since bias reduces trust, if we could get trust low enough by increasing bias, then algorithms could be similarly or more acceptable than humans, possibly eliminating bias tolerance.

STUDY 3: THE ROLE OF FAIRNESS

In this study, we had two goals. First, we wanted to learn why people disregard human bias and trust human decisions more than algorithmic decisions. Based on the affective human-likeness literature, we hypothesized that emotionality could explain this effect. Research has shown that machines such as robots are viewed as lacking human nature abilities, which are emotional (Haslam 2006; Loughnan and Haslam 2007). Affective human-likeness (which are the human capabilities that are affective/emotional) increases the use of algorithms in certain tasks (Castelo et al. 2019). Human-likeness can also indicate being biased. Our second goal was to determine whether unfairness influences perceptions of bias similarly for humans and algorithms. We aimed to create a scenario where bias is so strong that it diminishes the higher levels of trust in humans. Since bias reduces trust, if we could get trust low enough through an unfair outcome, then algorithms could be judged similarly or even more trustworthy relative to humans and bias tolerance might disappear. Accordingly, we manipulated choice type, outcome fairness, and outcome valence in the context of a loan application decision.

Method

Study 3 (*N* = 1280, MTurk) had a 2 (Decision type: Human Decision, Algorithmic Decision) x 2 (Outcome Fairness: Fair Outcome, Unfair Outcome) x 2 (Outcome Valence: Loan Approval, Loan Rejection) between-subjects design. Table 8 exhibits our experimental stimuli.

After reading their assigned stimuli, participants rated how satisfied they are with the bank agent's/algorithm's decision (1 = "Very dissatisfied", 7 = "Very satisfied"), how much they trust this bank agent's/algorithm's ability to decide on this loan application (1 = "Strongly distrust", 7 = "Strongly trust"), how likely they are to use this bank's services in the future (1 = "Very unlikely", 7 = "Very likely"), how biased they think this bank agent/algorithm is (1 = "Very unbiased", 7 = "Very biased"), and the extent to which they would describe the bank agent/algorithm as emotional (1 = "Not at all", 5 = "A great deal").

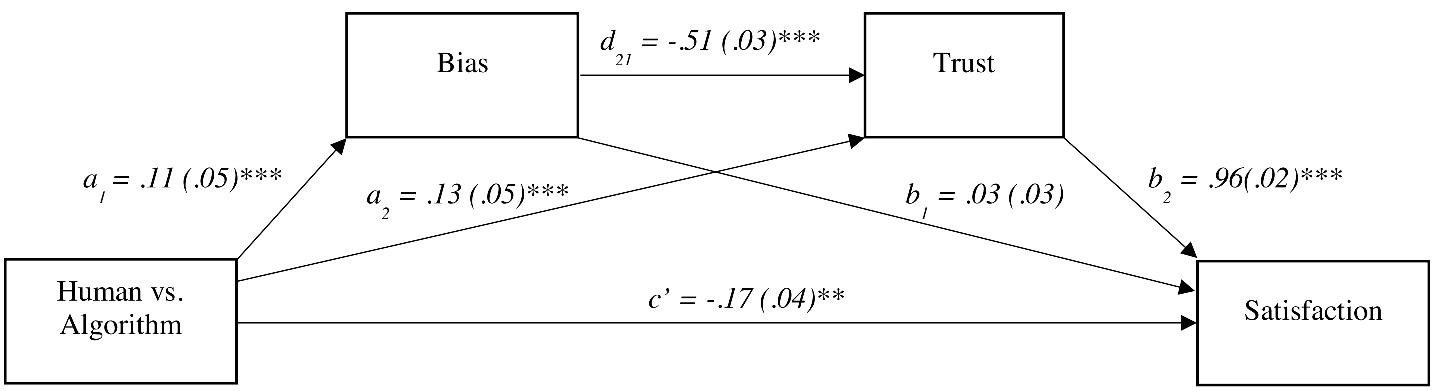
TABLE 8 (CHAPTER 2). STUDY 3: EXPERIMENTAL STIMULI

	Fair + Outcome	**Unfair + Outcome**	**Fair - Outcome**	**Unfair - Outcome**
Human Choice	Imagine that you applied your bank for a loan and that your loan application was approved. **The bank agent** evaluated your loan application and decided to **approve** it. This was a **fair outcome**, since the agent approved your application as they had determined that your ability to repay was higher than other applicants.	Imagine that you applied your bank for a loan and that your loan application was approved. **The bank agent** evaluated your loan application and decided to **approve** it. This was an **unfair outcome**, since the agent approved your application although they had determined that your ability to repay was lower than other applicants.	Imagine that you applied your bank for a loan and that your loan application was denied. **The bank agent** evaluated your loan application and decided to **reject** it. This was a **fair outcome**, since the agent rejected your application as they had determined that your ability to repay was lower than other applicants.	Imagine that you applied your bank for a loan and that your loan application was denied. **The bank agent** evaluated your loan application and decided to **reject** it. This was an **unfair outcome**, since the agent rejected your application although they had determined that your ability to repay was higher than other applicants.
Algorithmic Choice	Imagine that you applied your bank for a loan and that your loan application was approved. **The bank's algorithm** evaluated your loan application and decided to **approve** it. This was a **fair outcome**, since the algorithm approved your application as it had determined that your ability to repay was higher than other applicants.	Imagine that you applied your bank for a loan and that your loan application was approved. **The bank's algorithm** evaluated your loan application and decided to **approve** it. This was an **unfair outcome**, since the algorithm approved your application although it had determined that your ability to repay was lower than other applicants.	Imagine that you applied your bank for a loan and that your loan application was denied. **The bank's algorithm** evaluated your loan application and decided to **reject** it. This was a **fair outcome**, since the algorithm rejected your application as it had determined that your ability to repay was lower than other applicants.	Imagine that you applied your bank for a loan and that your loan application was denied. **The bank's algorithm** evaluated your loan application and decided to **reject** it. This was an **unfair outcome**, since the algorithm rejected your application although it had determined that your ability to repay was higher than other applicants.

Results

Planned comparisons indicated that human (vs. algorithmic) decisions were more trustworthy (F(1, 1278) = 25.52, p < .001), emotional (F(1, 1278) = 175.90, p < .001), biased (F(1, 1278) = 4.99, p = .026), and led to more future bank usage (F(1, 1278) = 6.604, p = .010). There was no significant difference on satisfaction F(1, 1278) = 1.656, p = .198). Our serial mediation analysis (Figure 11) replicated Study 1 bias tolerance findings: even though humans were perceived as more biased, they were still more trustworthy and their decisions lead to higher satisfaction levels.

FIGURE 11 (CHAPTER 2). STUDY 3: SERIAL MEDIATION ANALYSIS RESULTS

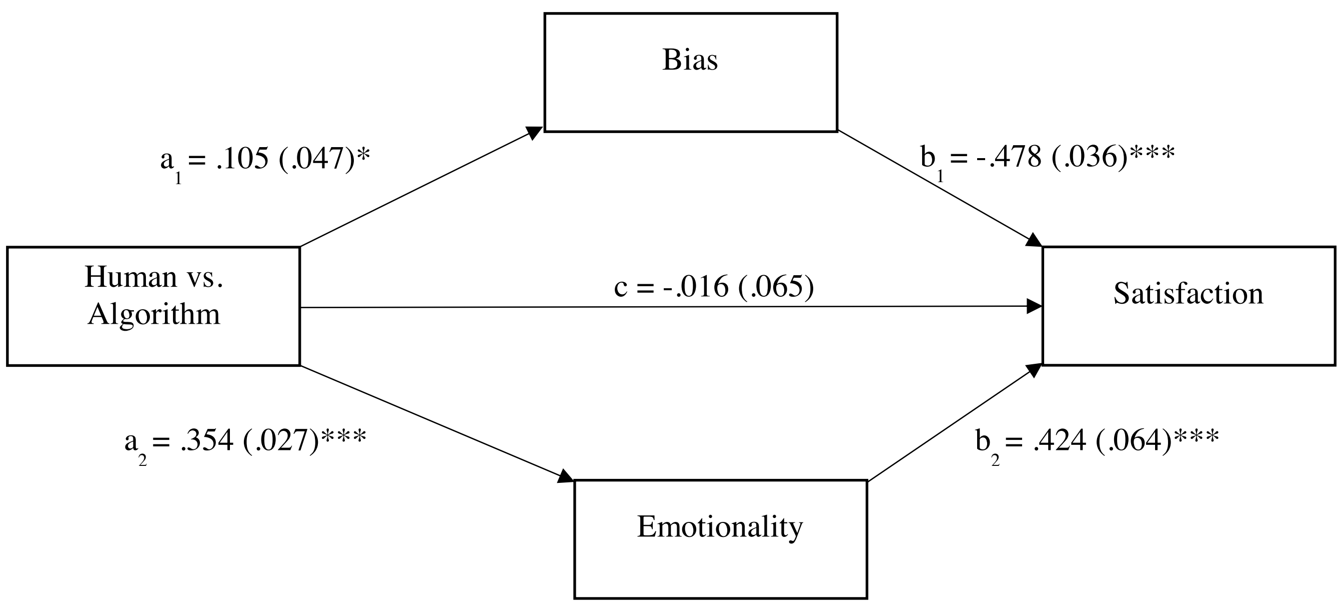


Note: * p < .05, *** p < .001. Parameter estimates and standard errors (in parentheses) are indicated. Indirect effect ($a_1 * d_{21} * b_2$) = -.051**, 95% CI = [-.099, -.006]

Additionally, we confirmed that the bias tolerance existed even in unfair decisions, indirect effect = -.129, 95% CI = [-.193, -.076]. Importantly, bias tolerance existed even in unfair decisions with negative outcomes: indirect effect = -.052, 95% CI = [-.099, -.024].

We wanted to investigate why bias neglect was such a robust enough effect that it even held for negative unfair decisions. A parallel mediation analysis with 10,000 bootstrapped samples revealed that emotionality helped explain why human (vs. algorithmic) decisions were more trustworthy, despite being acknowledged as more biased (Figure 12). The contrast between the indirect effect of bias and emotionality was significant: $\text{Bias}_{IDE} - \text{Emotionality}_{IDE} = -.200$, SE = .035, $p < .001$. The indirect effects were different whereby emotionality ($\text{Emotionality}_{IDE} = .150$, SE = .025, $p < .001$) mediated the relationship between choice type and satisfaction more strongly than bias ($\text{Bias}_{IDE} = -.050$, SE = .023, $p = .031$).

FIGURE 12 (CHAPTER 2). STUDY 3: EMOTIONALITY (VS. BIAS) IS A BETTER EXPLANATION FOR CHOICE TYPE-SATISFACTION RELATIONSHIP



Note: * $p < .05$, *** $p < .001$. $\text{Bias}_{IDE} = -.050$, SE = .023, $p = .03$. $\text{Emotionality}_{IDE} = .150$, SE = .025, $p < .001$. $\text{Bias}_{IDE} - \text{Emotionality}_{IDE} = -.200$, SE = .035, $p < .001$.

Discussion

In study 3, we find that even when the decision is explicitly unfair, people still discount human bias and trust human decisions more. The positive influence of emotionality on trust was greater than the negative effect of bias; humans were trusted more because they were more emotional, leading participants to disregard the negative influence of bias. Therefore, people might generally find human decision makers more biased than algorithmic decision makers but this does not offset the positive effect of human emotionality on trust. In study 4, we wanted to see if bias tolerance would exist in a more machine-like task (data handling) where there is little room for human emotionality and bias than in a more human-like task (decision making) where there is more room for human emotionality and bias.

STUDY 4: DECISION MAKING VERSUS DATA HANDLING

Study 4 broke down the two stages of receiving a decision in the digital world: data handling and decision-making. We aimed to investigate whether there is bias tolerance in both stages. We hypothesized that bias tolerance would not exist in the data handling stage since human emotionality and bias would be more irrelevant in a machine-like task compared to a more human-like task like decision-making. As the difference between human and algorithmic emotionality and bias is expected to be smaller in data handling (vs. decision making), we predicted that we would not see bias tolerance. In contrast, humans and algorithms would have more divergent bias and emotionality in decision making. Therefore, we predicted that we would observe bias tolerance in the decision-making stage, as we did in previous studies.

Method

Six hundred and forty-one MTurk participants were randomly assigned in a 2 (Data handler: Human, Algorithm) x 2 (Decision-maker: Human, Algorithm) between-subjects design (Table 9). Participants read that their personal data for a loan application was handled either by a human or an algorithm and decided by a human or an algorithm.

After reading whether their data was handled by a human or an algorithm, participants indicated how satisfied they are with how their data was handled (1 = “Very dissatisfied”, 7 = “Very satisfied”), how biased they think the algorithm was in handling their data (1 = “Very unbiased”, 7 = “Very biased”), how much they trust this advisor/algorithm in handling their data (1 = “Strongly distrust”, 7 = “Strongly trust”), and how emotional they think the advisor/algorithm was in handling their data (1 = “Very unemotional”, 7 = “Very emotional”).

Next, they read whether a financial advisor or an algorithm determined their loan application. They rated how satisfied they are with how their loan application was decided (1 = “Very dissatisfied”, 7 = “Very satisfied”), how biased they think the advisor/algorithm was in making this loan decision (1 = “Very unbiased”, 7 = “Very biased”), how much they trust this advisor/algorithm in deciding on this loan application (1 = “Strongly distrust”, 7 = “Strongly trust”), and how emotional they think advisor/algorithm was in making this decision (1 = “Very unemotional”, 7 = “Very emotional”).

TABLE 9 (CHAPTER 2). STUDY 4: EXPERIMENTAL STIMULI

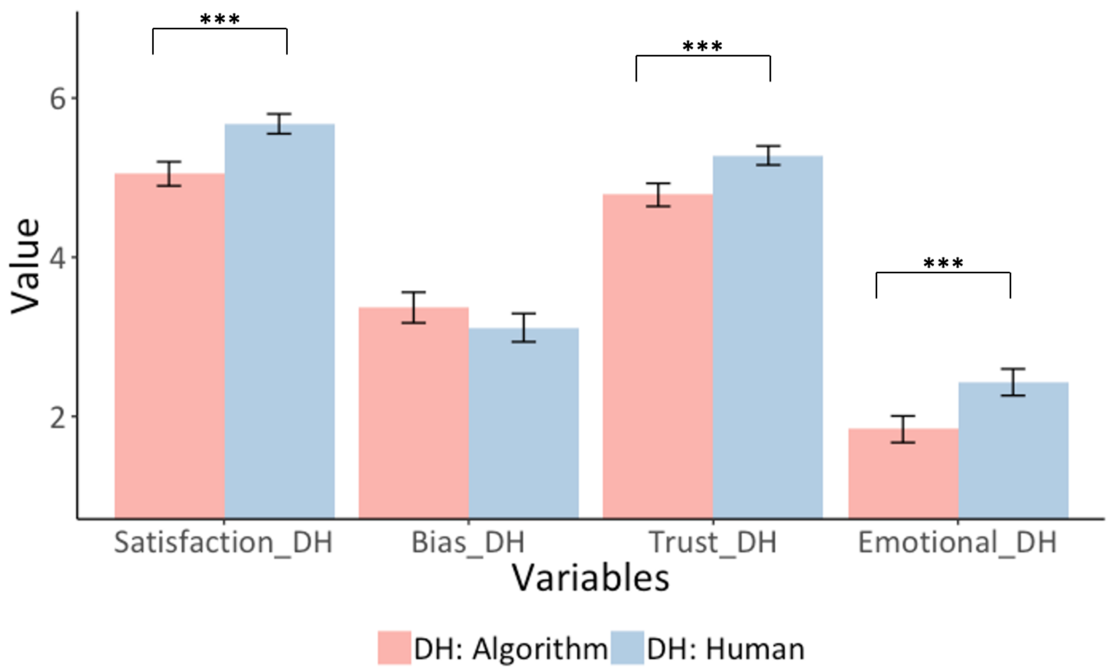
	Data Handler: Human	**Data Handler: Algorithm**
Decision Maker: Human	Imagine that you seek to obtain a loan and you apply a bank. In order to determine your loan allocation, **bank's financial advisor handled your data** related to your debt-to-income ratio, record of paying your bills on time, the value of your assets. **The advisor collected, organized, and analyzed your data**. After the advisor collected, organized, and analyzed your data, **your bank relied on the advisor to determine whether to accept or decline your loan application**. The advisor determined to decline your loan application.	Imagine that you seek to obtain a loan and you apply a bank. In order to determine your loan allocation, **bank's algorithm handled your data** related to your debt-to-income ratio, your record of paying your bills on time, the value of your assets. **The algorithm collected, organized, and analyzed your data**. After the algorithm collected, organized, and analyzed your data, **your bank relied on a financial advisor to determine whether to accept or decline your loan application.** The advisor determined to decline your loan application.
Decision Maker: Algorithm	Imagine that you seek to obtain a loan and you apply a bank. In order to determine your loan allocation, **bank's financial advisor handled your data** related to your debt-to-income ratio, record of paying your bills on time, the value of your assets. **The advisor collected, organized, and analyzed your data**. After the advisor collected, organized, and analyzed your data, **your bank relied on an algorithm to determine whether to accept or decline your loan application**. The algorithm determined to decline your loan application.	Imagine that you seek to obtain a loan and you apply a bank. In order to determine your loan allocation, **bank's algorithm handled your data related to your debt-to-income ratio**, your record of paying your bills on time, the value of your assets. **The algorithm collected, organized, and analyzed your data**. After the algorithm collected, organized, and analyzed your data**, your bank relied on the algorithm to determine whether to accept or decline your loan application**. The algorithm determined to decline your loan application.

Results

We found that human (vs. algorithmic) data handlers were more satisfactory ($F(1, 639) = 39$, $p < .001$), trustworthy ($F(1, 639) = 26.46$, $p < .001$), and emotional ($F(1, 639) = 23.97$, $p <$

.001). Algorithmic and human bias did not differ significantly for data handling ($F(1, 639) = 3.542$, $p < .060$).

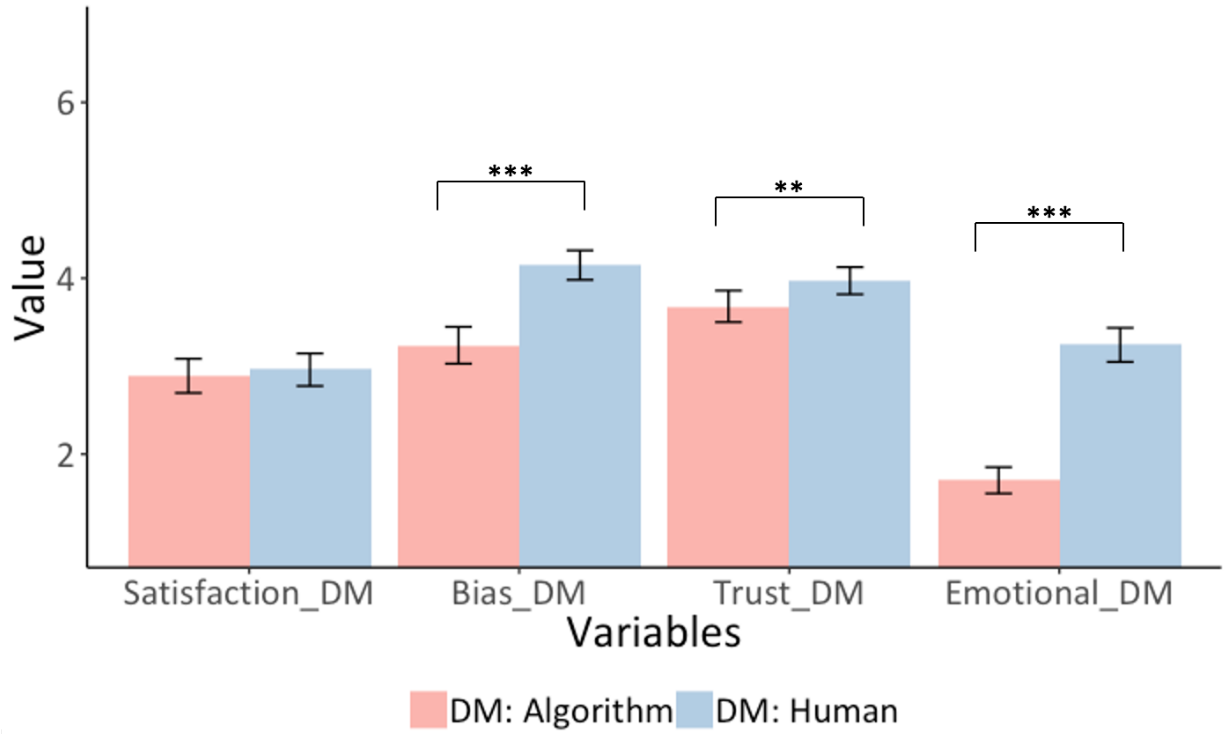
FIGURE 13 (CHAPTER 2). STUDY 4: DATA HANDLING RESULTS



Note: *** $p < .001$. Error bars represent ± standard error of the mean (SEM).

In contrast, in the decision-making stage, planned comparisons showed that humans were more biased ($F(1, 639) = 44.37$, $p < .001$), trustworthy ($F(1, 639) = 5.84$, $p < .016$), and emotional ($F(1, 639) = 151.60$, $p < .001$). Satisfaction wasn't significantly different between conditions ($F(1, 639) = .257$, $p = .613$).

FIGURE 14 (CHAPTER 2). STUDY 4: DECISION MAKING RESULTS



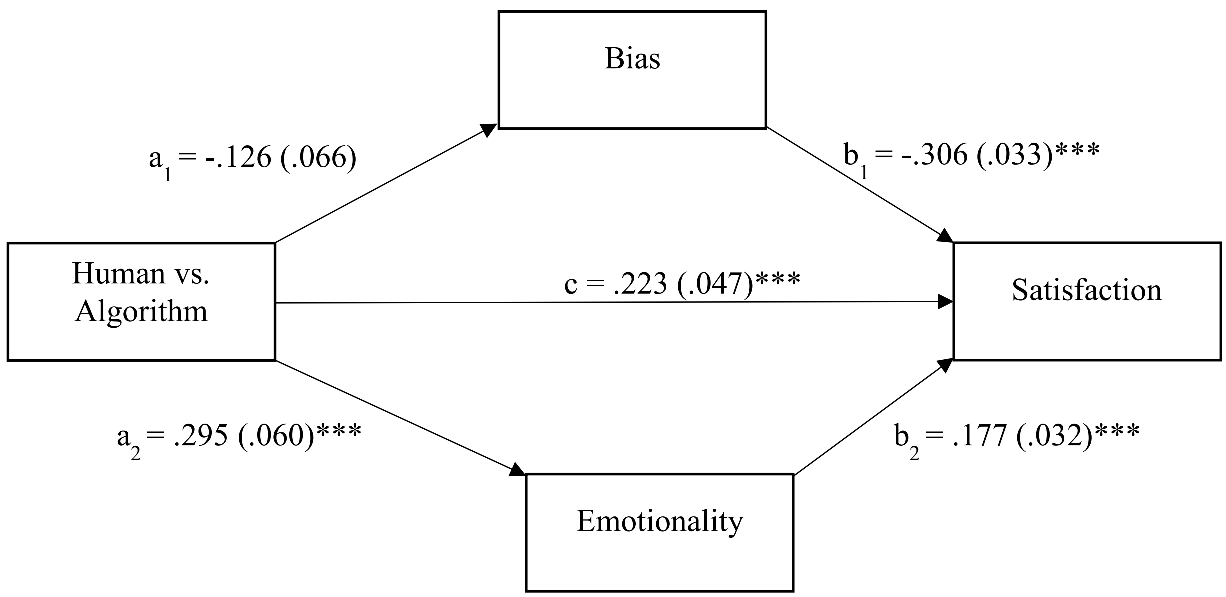
Note: ** $p < .01$, *** $p < .001$. Error bars represent ± standard error of the mean (SEM).

As in previous studies, human decision-makers were perceived as more biased, trustworthy, and satisfactory, and additional analysis shows that the higher human bias had no effect on satisfaction (B = .036, 95% CI = [-.034, .103]). Hence, while there is no apparent bias tolerance in the data handling stage (due to algorithms being perceived as equally or even marginally more biased than humans), bias tolerance did appear in the decision-making stage, where bias does not negatively impact satisfaction.

A parallel mediation analysis with 10,000 bootstrapped samples using the mediation effects of bias and emotionality for the relationship between choice type and satisfaction in the data handling stage was conducted (Figure 15). The contrast between the indirect effect of bias and emotionality was not significant: $\text{Bias}_{\text{IDE}} - \text{Emotionality}_{\text{IDE}} = -.014$, SE = .027, $p = .611$.

The indirect effects were not different where emotionality ($\text{Emotionality}_{IDE} = .052$, SE = .014, $p < .001$) and bias ($\text{Bias}_{IDE} = .039$, SE = .021, $p = .067$) mediated the relationship between choice type and satisfaction similarly.

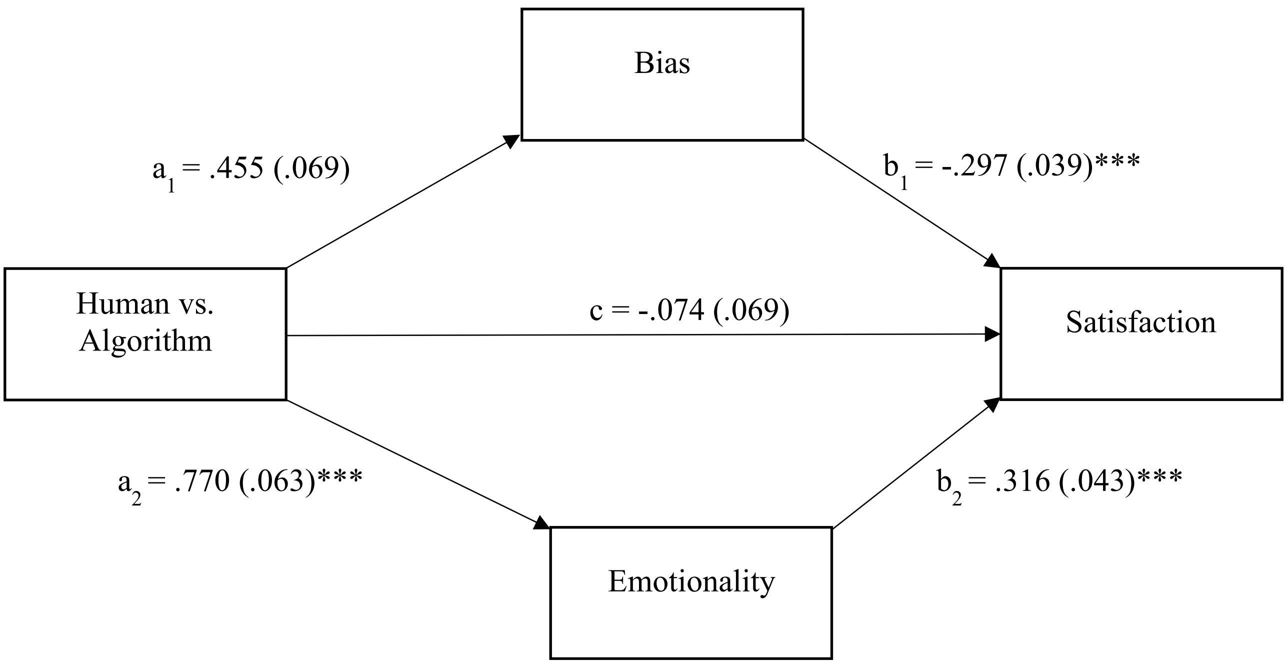
FIGURE 15 (CHAPTER 2). STUDY 4: EMOTIONALITY AND BIAS SIMILARLY EXPLAIN CHOICE TYPE-SATISFACTION RELATIONSHIP IN DATA HANDLING



Note: *** $p < .001$. $\text{Bias}_{IDE} = .039$, SE = .021, $p = .067$. $\text{Emotionality}_{IDE} = .052$, SE = .014, $p < .001$. $\text{Bias}_{IDE} - \text{Emotionality}_{IDE} = -.014$, SE = .027, $p = .611$.

In contrast, the same mediation analysis in the decision making stage (Figure 16) showed that the contrast between the indirect effect of bias and emotionality was significant: $\text{Bias}_{IDE} - \text{Emotionality}_{IDE} = -.379$, SE = .052, $p < .001$. The indirect effects were different such that emotionality ($\text{Emotionality}_{IDE} = .244$, SE = .039, $p < .001$) mediated the relationship between choice type and satisfaction more than bias (Bias_{IDE} = -.135, SE = .027, p < .001), just like in study 3.

FIGURE 16 (CHAPTER 2). STUDY 4: EMOTIONALITY IS A BETTER EXPLANATION THAN BIAS IN DECISION MAKING



Note: *** p < .001. Bias_{IDE} = -.135, SE = .027, p < .001. $\text{Emotionality}_{IDE}$ = .244, SE = .039, p < .001. Bias_{IDE} – $\text{Emotionality}_{IDE}$ = -.379, SE = .052, p < .001.

Discussion

In study 4, we see that the gap between human and algorithms on emotionality and bias was smaller in data handling (vs. decision making). The parallel mediations showed that when this difference is smaller, in the case of data handling, emotionality and bias similarly mediate the relationship between choice type and satisfaction. Humans are still preferred in data handling although this does not seem driven by differences in bias and emotionality. When the gap between humans and algorithms in emotionality and bias is bigger, in the case of decision making, the relationship between choice type and satisfaction is more strongly explained by emotionality than bias. Therefore, primarily emotionality, not bias, drives satisfaction in decision making as seen in studies 3 and 4. This suggests that while people tolerate bias, they reward emotionality in decision making.

GENERAL DISCUSSION

Overall, we find that individuals perceive human decision makers as often more biased than algorithmic decision makers, and yet this bias does not negatively impact trust and satisfaction for humans relative to algorithms. Our studies regularly show scenarios where bias might be advantageous, such as subjective or experiential decisions. In those scenarios, we find that although humans are usually perceived as more biased, they are still more trusted and satisfactory than algorithms. This was true even when we looked at extreme decisions (unfair, negative outcomes). In none of our studies do we see human bias as a reason to switch to algorithmic decisions. We also included scenarios where we thought strong examples of human bias (such as unfair decisions) might be penalized, and yet bias tolerance persisted in those scenarios too. We also, in study 4, included scenarios where human bias and emotionality might be minimized, such as data handling tasks, and found that while bias did not differ by human versus algorithm, satisfaction still remains higher for humans.

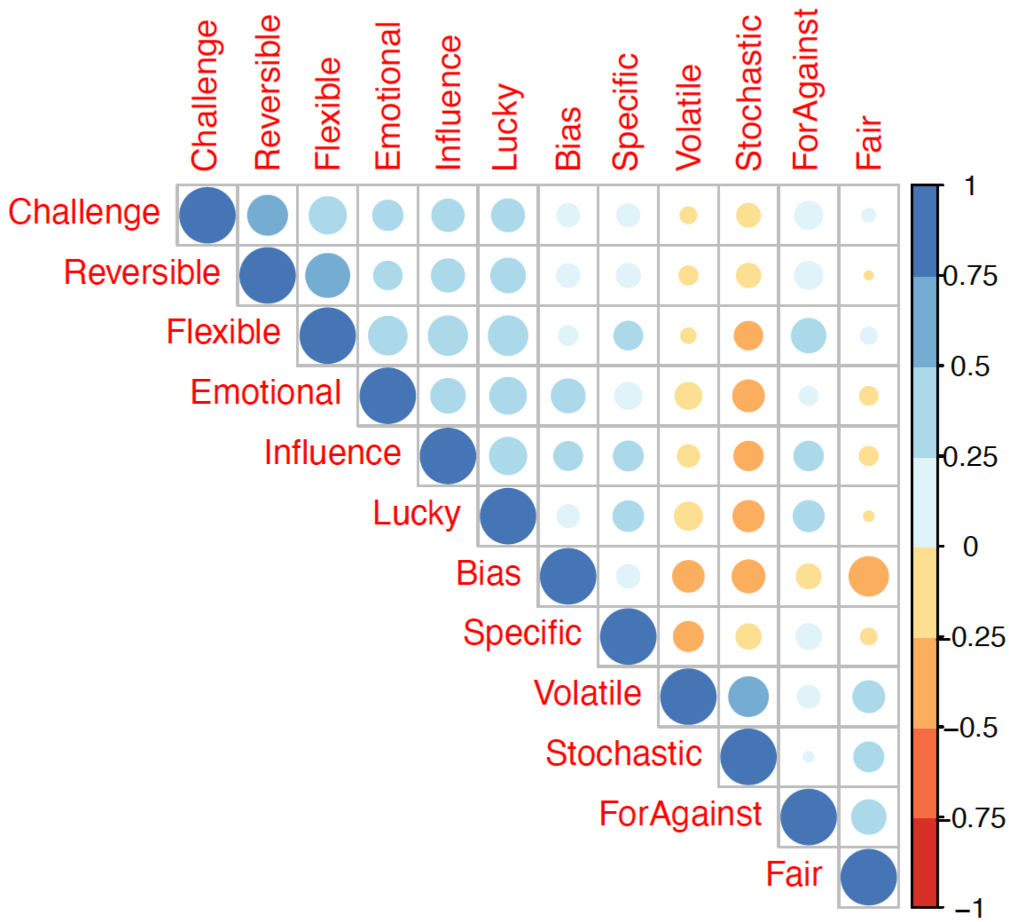
We aimed to understand what elements of bias might affect trust and satisfaction. Based on the affective human likeness literature, we thought emotionality could be a promising measure. We found that bias and emotionality are highly related. In studies 3 and 4 we found that emotionality, not bias, is a better explanation for why consumers are more satisfied with human (vs. algorithmic) decisions. Thus, it is not that bias does not matter, it is more that emotionality (which is correlated with bias) matters more. Our results indicate that while people are tolerating bias, they are heavily rewarding emotionality in human decision making.

Future Directions

What remain to be explored are the other factors that could contribute to the perceptions of bias, besides emotionality. In future work we should be accounting for the alternate mediators to understand why bias might not be predicting satisfaction. There are potential suppressing mediators that could make the relationship between bias and satisfaction null, even though we have a main effect of choice type on satisfaction. This opens future directions addressing many of the variables we tested in an ancillary study (Figure 17).

In this supplemental study where we investigated determinants of bias, we show how 11 variables correlate with bias. We find 4 measures that correlate heavily with bias: influenceability (whether the decision maker is viewed as being influenceable by the consumer), stochasticity (whether the decision maker would arrive at the same decision outcome if they made this same decision again tomorrow), volatility (how volatile they think the decision maker is) and fairness (whether they think they will be treated fairly by this decision maker). This can tell us about what else, besides emotionality, might be wrapped up in bias perceptions that might be causing the relationship between bias and satisfaction to be surprisingly null. Future work needs to be more precise about measuring all the variables that are correlated with perceptions of bias.

FIGURE 17 (SUPPLEMENTAL STUDY). CORRELATION PLOT: DETERMINANTS OF BIAS



Future research should continue to dive into question of what elements of bias are people willing to ignore and whether there any types of bias (discriminatory) that directly reduce trust. We can investigate whether we can reverse bias tolerance in other ways than we have in the studies thus far (unfair, negative outcomes and data handling). There could be circumstances where bias decreases satisfaction for humans and lack of bias becomes highly favorable for algorithms. In particular, studies could be conducted with marginalized populations (e.g., Asian

Americans in the context of college admissions) to see how perceptions of bias can change when the bias is perceived as racially motivated and potentially more impactful if it comes from humans (vs. algorithms).

It is also notable that in study 4, trust and satisfaction remain higher for human data handlers even when bias is perceived as similar for humans and algorithms. This suggests that there remain other drivers of satisfaction that penalize algorithms, such as general algorithm aversion (Dietvorst et al. 2015, 2018). Additional research should continue to explore the multiple possible mediators that affect satisfaction for algorithms.

Conclusion

Across four studies we studied “bias tolerance”: although algorithms were viewed as less biased than humans, human decisions were more trustworthy and satisfactory. This is true even in unfair decisions with negative outcomes. We find that bias tolerance occurs since human emotionality is highly desired, and this desire for emotionality outweighs the effects of higher perceived bias.

APPENDIX A: SUPPLEMENTAL MATERIAL FOR CHAPTER 1

SECTION 1: STUDY 1A ADDITIONAL ANALYSES

The linear regressions looking at the effect of choice type on satisfaction, trust, and process transparency controlling for preference for autonomy, showed that SC had higher satisfaction, trust, and process transparency than AC (contrast codes: Self Choice = 1, Algorithmic Choice = -1). The regression with the satisfaction DV demonstrated a significant effect of choice type (B = .129, SE = .055, $t(291) = 2.34$, $p < .001$), after controlling for autonomy (B = .036, SE = .024, $t(291) = 1.51$, $p = .131$), where algorithmic (vs. self) choice had lower satisfaction. The regression with trust as the dependent variable revealed a significant effect of choice type (B = .417, SE = .073, $t(291) = 5.69$, $p < .001$) , after controlling for autonomy (B = .137, SE = .032, $t(291) = 4.31$, $p < .001$), where algorithmic (vs. self) choice was less trustworthy. The results for the regression with process transparency as the dependent variable showed a significant effect of choice type (B = .449, SE = .080, $t(291) = 5.60$, $p < .001$) , after controlling for autonomy (B = .098, SE = .035, $t(291) = 2.81$, $p = .005$), where algorithmic (vs. self) choice had lower process transparency.

A parallel mediation analysis using the mediation effects of process transparency and preference for autonomy for the relationship between choice type and trust was conducted. The contrast between the indirect effect of process transparency and preference for autonomy was not significant: $\text{Process Transparency}_{IDE} - \text{Autonomy}_{IDE} = .014$, SE = .054, $p = .802$. The indirect effects were not different such that process transparency ($\text{Process Transparency}_{IDE} = .106$, SE = .033, $p = .002$) and preference for autonomy ($\text{Autonomy}_{IDE} = .092$, SE = .036, $p = .010$) similarly mediated the relationship between choice type and trust.

SECTION 2: STUDY 1B ADDITIONAL ANALYSES

We found that controlling for uniqueness neglect, SC had higher process transparency, trust, and decision comfort than AC (contrast codes: Self Choice = 1, Algorithmic Choice = -1). The linear regression with process transparency as the dependent variable revealed a significant effect of choice type (B = .622, *SE* = .113, $t(293) = 9.80$, $p < .001$) , after controlling for uniqueness neglect (B = .166, *SE* = .074, $t(293) = 2.24$, $p = .026$), where AC (vs. SC) had lower process transparency. The regression with trust as the DV showed a significant effect of choice type (B = .329, *SE* = .093, $t(293) = 3.53$, $p < .001$), after controlling for uniqueness neglect (B = -.021, *SE* = .061, $t(293) = -.34$, $p = .738$), where AC (vs. SC) choice was less trustworthy. The same was replicated with comfort as the DV: effect of choice type was significant (B = .206, *SE* = .082, $t(293) = 2.51$, $p = .013$), controlling for uniqueness neglect (B = .034, *SE* = .054, $t(293) = .63$, $p = .528$).

We conducted a parallel mediation analysis using the mediation effects of process transparency and uniqueness neglect for the relationship between choice type and trust. The contrast between the indirect effect of process transparency and uniqueness neglect was significant: Process $\text{Transparency}_{\text{IDE}}$ – Uniqueness $\text{Neglect}_{\text{IDE}}$ = .189, SE = .057, $p < .001$. The indirect effects were different whereby uniqueness neglect did not mediate the relationship between choice type and trust (Uniqueness $\text{Neglect}_{\text{IDE}}$ = .044, SE = .029, $p = .128$), while process transparency did (Process $\text{Transparency}_{\text{IDE}}$ = .233, SE = .050, $p < .001$). The same was replicated with comfort as the DV.

We also measured likelihood of betting all $25 on the chosen color after the primary DVs (1 = "Very unlikely", 7 = "Very likely"). Self and algorithmic choice participants did not differ in their likelihood of betting ($M_{\text{SC}} = 4.16$, $M_{\text{AC}} = 4.06$, $F(1, 294) = .159$, $p = .691$). Additionally, our exploratory analysis on the amount of the bet placed before the our comfort DV showed no difference between SC and AC (M_{SC} = 15.47, M_{AC} = 15.85, $F(1, 294)$ = .162, p = .688).

SECTION 3: STUDY 2 ADDITIONAL ANALYSES

	Label	Estimation	SE	p-value
SC vs. AC → Transparency	a	.76	.08	< .001
Transparency → Trust	b	.54	.05	< .001
SC vs. AC → Trust	c	.43	.10	< .001
Indirect Effect	a*b	.41	.05	< .001

	Label	Estimation	SE	p-value
SC vs. OC → Transparency	a	.21	.09	.016
Transparency → Trust	b	.60	.04	< .001
SC vs. OC → Trust	c	.37	.08	< .001
Indirect Effect	a*b	.13	.05	.017

SECTION 3: PRETEST

Digital domain descriptions:

Netflix is a streaming service that has a wide variety of TV shows, movies, documentaries, and more. You can either pick which show to watch yourself or the Netflix algorithm can choose a show for you.

Spotify is a digital music streaming service that gives you access to millions of songs and other content from artists all over the world. You can either pick which song to listen to yourself or the Spotify algorithm can choose a song for you.

Zipcar is a car-sharing service in which members can rent a variety of cars from a nearby Zipcar parking lot. You can either choose which car to rent yourself or the Zipcar algorithm can choose a car for you.

In **Google,** you can either do a Google search and choose a link yourself among search results or you can click on I'm Feeling Lucky and Google's algorithm chooses a link for you among search results.

In a **vending machine**, you can either pick a water bottle of your choice or vending machine's algorithm can pick a water bottle for you.

Uber is a ride-hailing service that dispatches drivers to passengers on demand. You can either pick one of several nearby drivers or the Uber algorithm can pick one of those drivers for you.

Bird is an electric scooter-sharing service where you can either choose a nearby scooter for your ride or the Bird algorithm can choose a nearby scooter for your ride.

AirBnB is a space-sharing service where you can either choose a property for your stay yourself or the AirBnB algorithm can choose a property for you.

YouTube is a video-sharing platform where you can either choose a video to watch yourself or the YouTube algorithm can choose a video for you to watch.

When driving, **you** can either control the car in all aspects of driving or the **self-driving car algorithm** can control the car in all aspects via its automated driving system.

While using news aggregators, such as **Apple News**, you can either choose the news source you would like to read from yourself or read the news that the news aggregator algorithm has curated for you.

When you are looking to buy a black t-shirt on **Amazon**, you can either choose a t-shirt yourself or Amazon's algorithm can choose a t-shirt for you (Amazon's Choice).

Google Maps is a web-based mapping service in which you can either pick your own route or Google Maps can pick the route for you.

When doing **retirement planning**, you can either choose the retirement plan yourself or an algorithmic advisor can choose it for you.

Additional Notes on Pretest

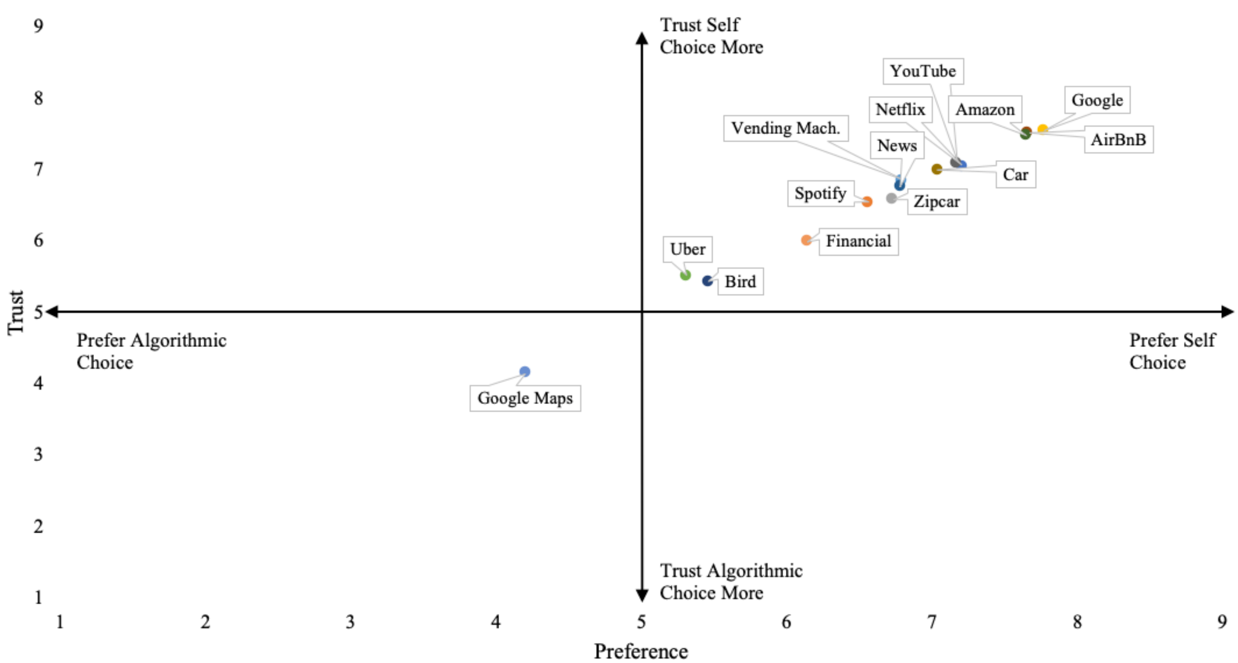
We counterbalanced the order of the domains. After reading a domain's description, they were asked whether they preferred choosing themselves or letting the algorithm choose for them in that particular domain on a 9-point Likert scale for (1 = "Definitely prefer algorithm choosing", 5 = "Indifferent between algorithm choosing or choosing myself", 9 = "Definitely prefer choosing myself"). They were also asked whether they trust their or the algorithm's ability to choose for them more in that domain (1 = "Definitely trust algorithm more than myself", 5 = "Equally trust algorithm and myself", 9 = "Definitely myself more than algorithm").

Google Maps was the only domain in which participants preferred and trusted algorithmic choice more ($M_{\text{Preference}} = 4.20$, $M_{\text{Trust}} = 4.16$, $r(248) = .92$, $p < .001$). Even in domains such as Google searches, where the "I'm Feeling Lucky" option provides a highly accurate algorithmic choice, people strongly preferred and trusted selecting an option themselves ($M_{\text{Preference}} = 7.77$, $M_{\text{Trust}} = 7.56$, $r(248) = .86$, $p < .001$).

Pretest Preference and Trust Results

Domain	Preference	Trust	Pearson's *r* (Correlation between Preference and Trust)
Google Maps	4.20	4.16	.92
Uber	5.31	5.52	.83
Bird	5.46	5.44	.85
Financial	6.14	6.00	.85
Spotify	6.56	6.55	.85
Zipcar	6.73	6.59	.90
News	6.78	6.77	.86
Vending Machine	6.79	6.85	.85
Car	7.04	7.00	.89
YouTube	7.17	7.10	.80
Netflix	7.21	7.05	.82
Amazon	7.65	7.49	.83
AirBnB	7.66	7.52	.85
Google	7.77	7.56	.86

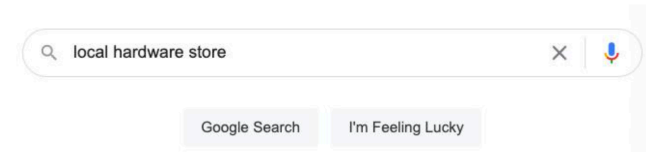
Preference For and Trust in Algorithmic (vs. Self) Choice



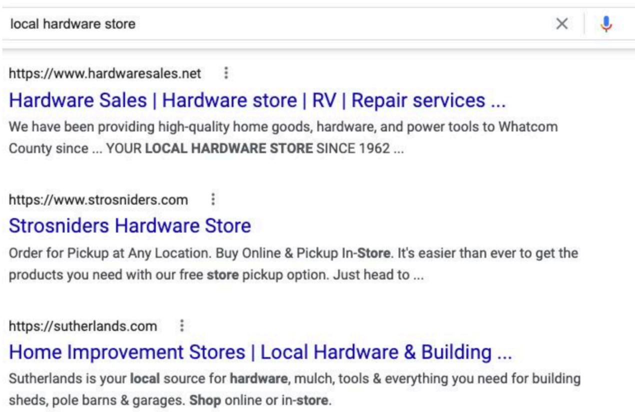
SECTION 4: STUDY 3 ADDITIONAL DETAILS

Stimuli

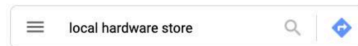
<table>
<tr><th></th><th>Google</th><th>Google Maps</th></tr>
<tr><td>SC</td><td>
You go onto Google to search for a local hardware store.



After clicking on Google Search, you see the 3 store options that Google found. You pick one of these options.

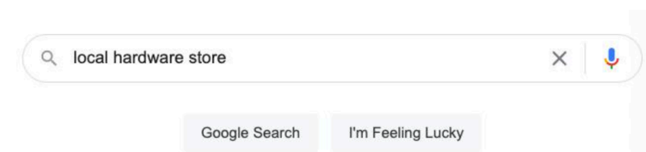


</td><td>
You go onto Google Maps to search for directions to a local hardware store.

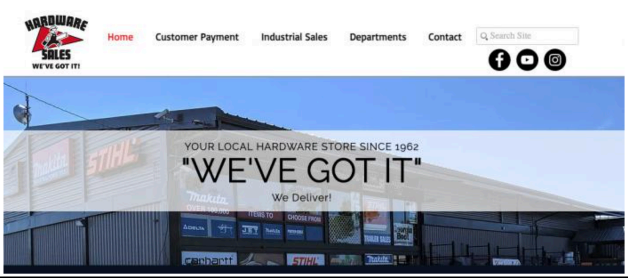


After typing in the name of the store, you see the 3 route options that Google Maps found. You pick one of these options and start your drive to the store.

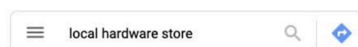
</td></tr>
<tr><td>AC</td><td>
You go onto Google to search for a local hardware store.



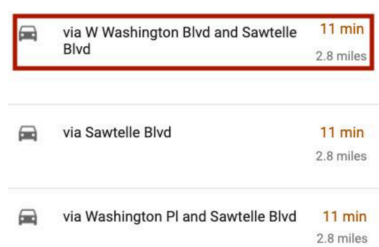
After clicking on "I'm Feeling Lucky", Google's algorithm chooses this hardware store among search results.



</td><td>
You go onto Google Maps to search for directions to a local hardware store.



After typing in the name of the store, you see the 3 route options that Google Maps found. The Google Maps algorithm picks one of these options and you start your drive to the store.



</td></tr>
</table>

Our confirmatory factor analysis found that trust and perceived accuracy are essentially the same construct. We performed a confirmatory factor analysis with trust and perceived accuracy loaded onto one factor and transparency loaded another factor. The results were the following: the comparative fit index (CFI) = 1.00, the Tucker-Lewis fit index (TLI) = 1.00, and the RMSEA < .01. These values suggested a good fit between the model and the observed data. Accordingly, we created a trust index with trust and perceived accuracy measures combined (Cronbach's alpha = .81). All our linear regressions and mediation analyses were replicated with trust, accuracy, and the trust index. The results did not change whether we used trust as the DV, perceived accuracy as the DV or combined measure of trust and perceived accuracy as the DV. We report trust results in the article because that is the DV we use in all the other studies.

SECTION 5: STUDY 4 ADDITIONAL ANALYSES

	Label	Estimation	SE	p-value
X vs. Baseline → Transparency	a_1	.69	.06	< .001
X vs. Baseline → Trust	a_2	.17	.06	.007
Transparency → Trust	d_{21}	.21	.03	< .001
X vs. Baseline → Satisfaction	c'	.26	.05	< .001
Transparency → Satisfaction	b_1	.06	.02	.006
Trust → Satisfaction	b_2	.58	.03	< .001
Indirect Effect	$a_1 * d_{21} * b_2$	.09	.02	< .001

	Label	Estimation	SE	p-value
UnX vs. Baseline → Transparency	a_1	-.24	.07	< .001
UnX vs. Baseline → Trust	a_2	-.05	.06	.348
Transparency → Trust	d_{21}	.24	.03	< .001
UnX vs. Baseline → Satisfaction	c'	.08	.04	.052
Transparency → Satisfaction	b_1	.10	.02	< .001
Trust → Satisfaction	b_2	.59	.03	< .001
Indirect Effect	$a_1 * d_{21} * b_2$	-.03	.01	.002

SECTION 6: SPOTIFY STUDY DETAILS

Method

We recruited three hundred eighty-seven Prolific participants (M_{age} = 30.1, SD = 9.96, 42.9% female) in exchange for monetary compensation. Participants were randomly assigned to one of two conditions: Input explainable, Input unexplainable. The following were the stimuli corresponding these conditions.

Explainable	Unexplainable
Imagine that you open your Spotify app. You see that the Spotify algorithm has created a "Missed Hits" playlist for you based on the popular songs in 2020 that you didn’t listen to, your favorite artists in 2020, and the genres you have enjoyed the most.	Imagine that you open your Spotify app. You see that the Spotify algorithm has created a "Missed Hits" playlist for you.

After reading their assigned scenario, all participants indicated how well they understand the process behind Spotify algorithm's choice of music in this playlist (1 = “Little to no understanding”, 4 = “Moderate understanding”, 7 = “Detailed and deep understanding”). Next, they rated how much they would trust Spotify algorithm's ability to decide on songs for them in this playlist (1 = “Strongly distrust”, 7 = “Strongly trust”) and how satisfied they would be with the Spotify algorithm's choice of music for them in this playlist in this scenario (1 = “Very unsatisfied”, 7 = “Very satisfied”).

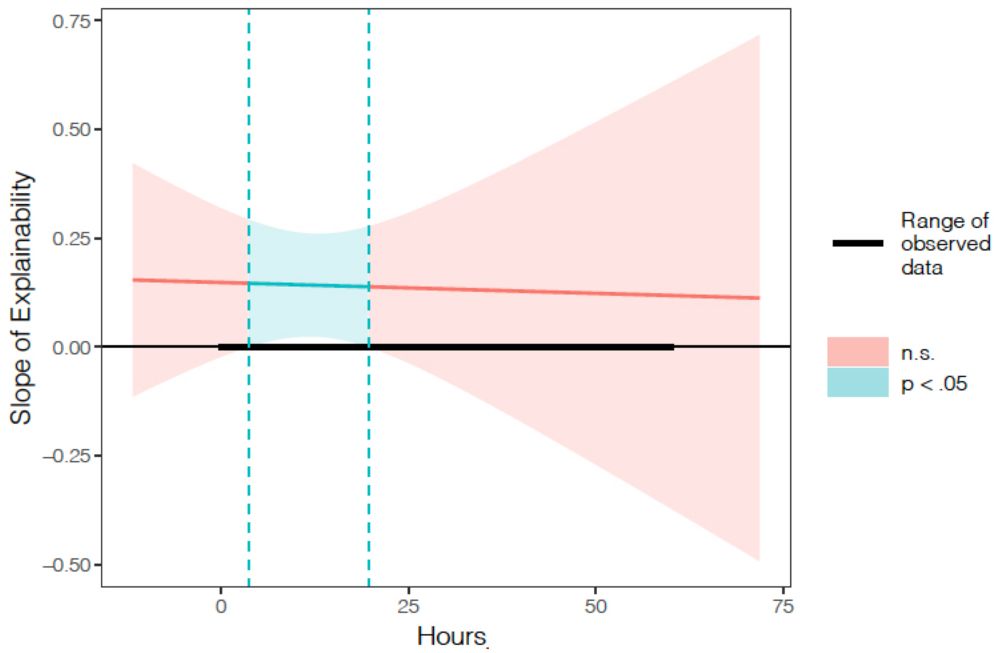
Results

Input explainable (vs. unexplainable) decisions were marginally significantly more process transparent ($M_X = 3.98$, $M_{UnX} = 3.75$, $F(1, 385) = 2.61$, $p = .107$) and significantly more satisfactory ($M_X = 4.81$, $M_{UnX} = 4.58$, $F(1, 385) = 3.96$, $p = .047$). There was no significant difference in trust ($M_X = 4.83$, $M_{UnX} = 4.72$, $F(1, 385) = .95$, $p = .331$), which might be due to the significant majority (%84.8) of our participants being Spotify users that already trust Spotify's ability to choose songs to create playlists.

Our floodlight analysis focused on our Spotify using participants[5] and tested whether Spotify usage (measured in the reported number of hours per week using Spotify) moderated the effect of input explainability on satisfaction. It revealed two Jonson-Neyman points: 3.92 and 19.62. In particular, input explainability had no effect on those who used Spotify more than 19.62 hours per week and those who used Spotify less than 3.92 hours per week. Our input explainability intervention was effective in increasing satisfaction for the 56.6% of our participants who used Spotify more than 3.92 hours, but less than 19.62 hours. Surprisingly, the mid-level users benefit from algorithm's input explainability the most and it had no significant effect on experienced or inexperienced Spotify users.

[5] Our floodlight analysis was conducted only with the participants who are Spotify users to make it more relevant. Furthermore, 3 participants who implausibly indicated using Spotify more than the number of hours per week and 3 standard deviations above the mean were removed from this analysis. Consequently, 325 participants were included in this analysis. Notwithstanding our focus on the Spotify users, the floodlight analysis with 384 participants look very similar, with JN points at 7.07 and 17.63.

FLOODLIGHT ANALYSIS: USAGE AS A MODERATOR OF INPUT EXPLAINABILITY'S EFFECT ON SATISFACTION



SECTION 7: FAIRNESS STUDY DETAILS

Results up to this point have shown the positive effect that an input explainability intervention can have on trust and satisfaction measures. Nevertheless, procedural fairness could moderate this effect. The fairness information conveyed could potentially affect the influence of explainability on trust and satisfaction such that if the procedures were explained to be unfair, then the explainability interventions could hurt trust and satisfaction. In contrast, if the procedural fairness is demonstrated in an explainability intervention, then trust and satisfaction could be boosted. Thus, it is crucial to investigate how procedural fairness is moderates the effect of an input explainability intervention.

Method

We recruited one thousand two hundred three participants from MTurk. The experiment involved a 2 (Choice type: Human Choice (HC), Algorithmic Choice (AC)) x 2 (Input Explainability: Explainable, Unexplainable) x 2 (Procedural Fairness: Fair, Unfair) between-subjects design, in which subjects were randomly assigned to one of eight conditions depicted in the following table.

	X, Fair	UnX, Fair	X, Unfair	UnX, Unfair
Human Choice	Imagine that you applied your bank for a loan and that your loan application was denied. The bank agent evaluated your loan application via a blind procedure. The agent based their assessment on: 5. Stability (how long you have been living at your current address and how long you have been in your current job) 6. Your debt-to-income ratio 7. The value of your assets 8. Your record of paying your bills on time and in their entirety	Imagine that you applied your bank for a loan and that your loan application was denied. The bank agent evaluated your loan application via a blind procedure.	Imagine that you applied your bank for a loan and that your loan application was denied. The bank agent evaluated your loan application via a procedure that could be biased. The agent based their assessment on: 1. Gender 2. Race 3. Age 4. Religion	Imagine that you applied your bank for a loan and that your loan application was denied. The bank agent evaluated your loan application via procedure that could be biased.
Algorithmic Choice	Imagine that you applied your bank for a loan and that your loan application was denied. The bank's algorithm evaluated your loan application via a blind procedure. The algorithm based its assessment on: 5. Stability (how long you have been living at your current address and how long you have been in your current job) 6. Your debt-to-income ratio 7. The value of your assets 8. Your record of paying your bills on time and in their entirety	Imagine that you applied your bank for a loan and that your loan application was denied. The bank's algorithm evaluated your loan application via a blind procedure.	Imagine that you applied your bank for a loan and that your loan application was denied. The bank's algorithm evaluated your loan application via a procedure that could be biased. The algorithm based its assessment on: 1. Gender 2. Race 3. Age 4. Religion	Imagine that you applied your bank for a loan and that your loan application was denied. The bank's algorithm evaluated your loan application via a procedure that could be biased.

After participants read their prompt, they wrote down as complete of an explanation of the loan decision as possible. Next, they completed the process transparency measure by indicating how well they understand why the bank agent (in the HC conditions) or the algorithm (in the AC conditions) denied their loan application on a 7-point scale (1 = "Little to no understanding", 4 = "Moderate understanding", 7 = "Detailed and deep understanding"). Then, they reported how dissatisfied they were with the decision on a 7-point scale (1 = "Very satisfied", 7 = "Very dissatisfied"), how much they trusted the bank agent (HC) or the bank algorithm (AC) to decide on this loan application on a 5-point scale (1 = "Not at all", 5 = "A great deal"), and how likely they are to use this bank's services in the future on a 7-point scale (1 = "Very unlikely", 7 = "Very likely"). Finally, they rated how biased they think the bank agent (HC) or the bank algorithm (AC) is (1 = "Very unbiased", 7 = "Very biased").

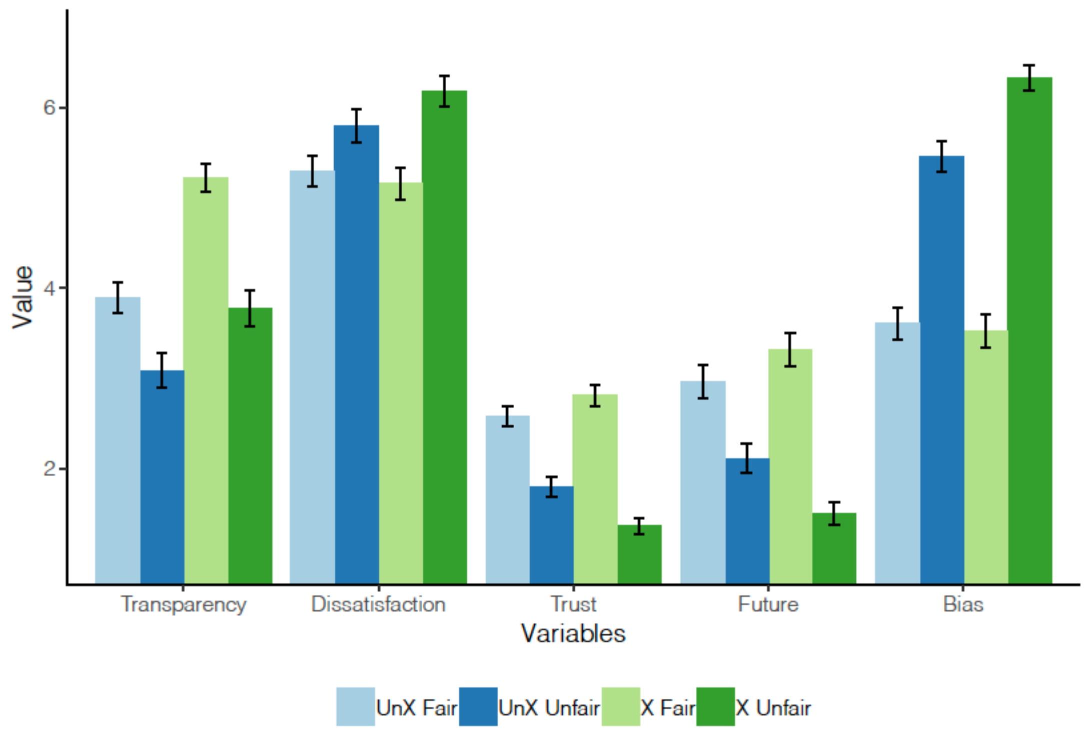
Results

In order to test our hypothesis that the effect of input explainability was moderated by fairness, we separately regressed process transparency (Model 1), dissatisfaction (Model 2), trust (Model 3), future use (Model 4), and bias (Model 5) on input explainability (Explainable = 1, Unexplainable = -1), process fairness (Fair = 1, Unfair = -1), and their interaction. There were significant interactions on all our dependent variables, as depicted in the table below.

FAIRNESS X EXPLAINABILITY INTERACTION RESULTS

	Dependent variable:				
	Transparency	Dissatisfaction	Trust	Future	Bias
	(1)	(2)	(3)	(4)	(5)
Fair.Unfair_contrast	0.563^{***} (0.046)	-0.381^{***} (0.045)	0.556^{***} (0.028)	0.665^{***} (0.043)	-1.163^{***} (0.043)
X.UnX_contrast	0.505^{***} (0.046)	0.062 (0.045)	-0.050^{*} (0.028)	−0.062 (0.043)	0.196^{***} (0.043)
Fair.Unfair_contrast:X.UnX_contrast	0.159^{***} (0.046)	-0.132^{***} (0.045)	0.168^{***} (0.028)	0.243^{***} (0.043)	-0.238^{***} (0.043)
Constant	3.995^{***} (0.046)	5.607^{***} (0.045)	2.136^{***} (0.028)	2.473^{***} (0.043)	4.730^{***} (0.043)
Observations	1,203	1,203	1,203	1,203	1,203
R^2	0.189	0.065	0.274	0.188	0.398
Adjusted R^2	0.187	0.063	0.273	0.186	0.397
Residual Std. Error (df = 1199)	1.597	1.555	0.954	1.487	1.489
F Statistic (df = 3; 1199)	93.133^{***}	27.929^{***}	151.192^{***}	92.777^{***}	264.378^{***}
Note:					*p<0.1; **p<0.05; ***p<0.01

FAIRNESS STUDY RESULTS



Explainability increased transparency more for fair than unfair decisions. Explainability increased dissatisfaction and bias for unfair decisions, while it did not change them for fair decisions. Explainability increased trust and future bank usage for fair but decreased them for unfair decisions. There were no 3-way interactions: the moderating effect of procedural fairness on input explainability was similar for human and algorithmic decisions.

Our moderated mediation analysis demonstrated that the indirect effect of input explainability through process transparency on dissatisfaction was stronger for unfair than fair decisions: estimated difference of ACME(Fair) – ACME(Unfair) was -0.077, 95% CI = [-.150, -.010].

REFERENCES

Alba, Joseph W and J Wesley Hutchinson (2000), "Knowledge Calibration: What Consumers Know and What They Think They Know," *Journal of Consumer Research*, 27(2), 123–56.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner (2016), "Machine Bias," *Propublica*, 1–26, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Angwin, Julia, Surya Mattu, and Jeff Larson (2015), "The Tiger Mom Tax: Asians Are Nearly Twice as Likely to Get a Higher Price from Princeton Review," *ProPublica*, https://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review.

Arkes, Hal R, Victoria A Shaffer, and Mitchell A Medow (2007), "Patients Derogate Physicians Who Use a Computer-Assisted Diagnostic Aid," *Medical Decision Making*, 27(2), 189–202.

Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan (2018), "The Moral Machine Experiment," *Nature*, 563(7729), 59–64.

Awad, Edmond, Sohan Dsouza, Iyad Rahwan, Azim Shariff, and Jean-François Bonnefon (2017), "Moral Machine," http://moralmachine.mit.edu/.

Barocas, Solon and Andrew Selbst (2016), "Big Data's Disparate Impact," *California Law Review*, 104(3), 671.

Bensinger, Greg and Tim Higgins (2018), "Uber Suspends Driverless-Car Program After Pedestrian Is Killed," *The Wall Street Journal*, https://www.wsj.com/articles/uber-suspends-driverless-car-program-after-pedestrian-is-killed-1521551002.

Blader, Steven L. and Ya Ru Chen (2011), "What Influences How Higher-Status People

Respond to Lower-Status Others? Effects of Procedural Fairness, Outcome Favorability, and Concerns about Status," *Organization Science*, 22(4), 1040–60.

Bonnefon, Jean Francois, Azim Shariff, and Iyad Rahwan (2016), "The Social Dilemma of Autonomous Vehicles," *Science*, 352(6293), 1573–76.

Botti, Simona, Kristina Orfali, and Sheena S Iyengar (2009), "Tragic Choices: Autonomy and Emotional Responses to Medical Decisions," *Journal of Consumer Research*, 36(3), 337–52.

Van Boven, Leaf and Thomas Gilovich (2003), "To Do or to Have? That Is the Question," *Journal of Personality and Social Psychology*, 1193–1202.

Brockner, Joel (2002), "Making Sense of Procedural Fairness : How High Procedural Fairness Can Reduce or Heighten the Influence of Outcome Favorability," *The Academy of Management Review*, 27(1), 58–76.

Brockner, Joel and Batia M Wiesenfeld (1996), "An Integrative Framework for Explaining Reactions to Decisions: Interactive Effects of Outcomes and Procedures," *Psychological Bulletin*, 120(2), 189–208.

Camerer, Colin (1981), "General Conditions for the Success of Bootstrapping Models," *Organizational Behavior and Human Performance*, 27(3), 411–22.

Carter, Travis J. and Thomas Gilovich (2010), "The Relative Relativity of Material and Experiential Purchases," *Journal of Personality and Social Psychology*, 98(1), 146–59.

Caruana, Rich, Yin Lou, Johannes Gehrke, Paul Koch, Marc Sturm, and Noémie Elhadad (2015), "Intelligible Models for Healthcare: Predicting Pneumonia Risk and Hospital 30-Day Readmission," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1721–30.

Castelo, Noah, Maarten W. Bos, and Donald R. Lehmann (2019), "Task-Dependent Algorithm Aversion," *Journal of Marketing Research*, 56(5), 809–25.

Choi, Jong Kyu and Yong Gu Ji (2015), "Investigating the Importance of Trust on Adopting an Autonomous Vehicle," *International Journal of Human-Computer Interaction*, 31(10), 692–702.

Colson, Eric, Brian Coffey, Tarek Rached, and Liz Cruz (2021), "Stitch Fix Algorithms Tour," *Stitch Fix*, https://algorithms-tour.stitchfix.com/#data-platform.

Conlon, Donald E. and Noel M. Murray (1996), "Customer Perceptions of Corporate Responses to Product Complaints: The Role of Explanations," *Academy of Management Journal*, 39(4), 1040–56.

Cowgill, Bo (2020), *Bias and Productivity in Humans and Algorithms: Theory and Evidence from Résumé Screening*, 29 Columbia Business School Research Paper Series.

Dai, Hengchen, Cindy Chan, and Cassie Mogilner (2020), "People Rely Less on Consumer Reviews for Experiential than Material Purchases," *Journal of Consumer Research*, 46(6), 1052–75.

Dastin, Jeffrey (2018), "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women - Reuters," *Reuters*, 5–9, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Dawes, Robyn M. (1979), "The Robust Beauty of Improper Linear Models in Decision Making," *American Psychologist*, 34(7), 571–82.

Dawes, Robyn M., David Faust, and Paul E. Meehl (1989), "Clinical versus Actuarial Judgment," *Science*, 243(4899), 1668–74.

Dawes, Robyn M and Bernard Corrigan (1974), "Linear Models in Decision Making," *Psychological Bulletin*, 81(2), 95–106.

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2015), "Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err," *Journal of Experimental Psychology: General*, 144(1), 114–26.

——— (2018), "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them," *Management Science*, 64(3), 1155–70.

Dunning, David (2011), "The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance," in *Advances in Experimental Social Psychology*, 247–96.

Dunning, David, Dale W. Griffin, James D. Milojkovic, and Lee Ross (1990), "The Overconfidence Effect in Social Prediction," *Journal of Personality and Social Psychology*, 58(4), 568–81.

Eastwood, Joseph, Brent Snook, and Kirk Luther (2012), "What People Want From Their Professionals: Attitudes Toward Decision-Making Strategies," *Journal of Behavioral Decision Making*, 25(5), 458–68.

Erel, Isil, Léa H. Stern, Chenhao Tan, and Michael S Weisbach (2017), *Selecting Directors Using Machine Learning*.

European Parliament and Council of the European Union (2016), *General Data Protection Regulation-Regulation (EU) 2016/679*, Official Journal of the European Union 1–88.

Friedman, Batya and Helen Nissenbaum (1996), "Bias in Computer Systems," *ACM Transactions on Information Systems (TOIS)*, 14(3), 330–47.

Gefen, David, Elena Karahanna, and Detmar W. Straub (2003), "Trust and TAM in Online Shopping: An Integrated Model," *MIS Quarterly*, 27(1), 51–90.

Grove, William M., David H. Zald, Boyd S. Lebow, Beth E. Snitz, and Chad Nelson (2000), "Clinical versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment*, 19–30.

Hannak, Aniko, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson (2014), "Measuring Price Discrimination and Steering on E-Commerce Web Sites," in *Proceedings of the ACM*, 305–18.

Harvey, Nigel and Ilan Fischer (1997), "Taking Advice Accepting Help, Improving Judgment," *Organizational Behavior and Human Decision Processes*, 70(2), 117–33.

Haslam, Nick (2006), "Dehumanization: An Integrative Review," *Personality and Social Psychology Review*, 10(3), 252–64.

Haslam, Nick, Yoshihisa Kashima, Stephen Loughnan, Junqi Shi, and Caterina Suitner (2008), "Subhuman, Inhuman, and Superhuman: Contrasting Humans with Nonhumans in Three Cultures," *Social Cognition*, 26(2), 248–58.

Hayes, Andrew F. (2013), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, Guilford Publications, New York, NY, US: Guilford Press.

Highhouse, Scott (2008), "Stubborn Reliance on Intuition and Subjectivity in Employee Selection," *Industrial and Organizational Psychology*, 1(3), 333–42.

Howard, John A. and Jagdish N. Sheth (1969), *The Theory of Buyer Behavior*, New York: John Wiley and Sons.

Joel R. Spiegel, Michael T. McKenna, Girish S. Lakshman, and Paul G. Nordstrom (2013), "Method and System for Anticipatory Package Shipping," *United States Patent*, 1–27.

Kahneman, Daniel, Olivier Sibony, and Cass R. Sunstein (2021), *Noise: A Flaw in Human*

Judgment, Little, Brown and Company.

Keil, Frank C. (2006), "Explanation and Understanding," *Annual Review of Psychology*, 57, 227–54.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2017), *Human Decisions and Machine Predictions*.

Korsgaard, M. A., D. M. Schweiger, and H. J. Sapienza (1995), "Building Commitment , Attachment , and Trust in Strategic Decision-Making Teams : The Role of Procedural Justice," *Academy of Management Journal*, 38(1), 60–84.

Lakkaraju, Himabindu, Ece Kamar, Rich Caruana, and Jure Leskovec (2017), "Interpretable & Explorable Approximations of Black Box Models," *arXiv preprint*.

Langer, Ellen J., Arthur Blank, and Benzion Chanowitz (1978), "The Mindlessness of Ostensibly Thoughtful Action: The Role of 'Placebic' Information in Interpersonal Interaction.," *Journal of Personality and Social Psychology*, 36(6), 635–42.

Lee, John D. and Neville Moray (1992), "Trust, Control Strategies and Allocation of Function in Human-Machine Systems," *Ergonomics*, 35(10), 1243–70.

Lee, John D. and Katrina A. See (2004), "Trust in Automation: Designing for Appropriate Reliance," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80.

Lim, Joa Sang and Marcus O'Connor (1995), "Judgemental Adjustment of Initial Forecasts: Its Effectiveness and Biases," *Journal of Behavioral Decision Making*, 8(3), 149–68.

Logg, Jennifer M., Julia A. Minson, and Don A. Moore (2019), "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment," *Organizational Behavior and Human Decision Processes*, 151, 90–103.

Lombrozo, Tania (2007), "Simplicity and Probability in Causal Explanation," *Cognitive Psychology*, 55(3), 232–57.

Longoni, Chiara, Andrea Bonezzi, and Carey K. Morewedge (2019), "Resistance to Medical Artificial Intelligence," *Journal of Consumer Research*, 46(4), 629–50.

Loughnan, Stephen and Nick Haslam (2007), "Animals and Androids: Implicit Associations between Social Categories and Nonhumans," *Psychological Science*, 18(2), 116–21.

Matsakis, Louise (2019), "What Does 'Amazon's Choice' Actually Mean?," *WIRED*, https://www.wired.com/story/what-does-amazons-choice-mean/.

Mattioli, Dana (2012), "On Orbitz, Mac Users Steered to Pricier Hotels," *Wall Street Journal*, https://www.wsj.com/articles/SB10001424052702304458604577488822667325882.

Mayer, Roger C., James H. Davis, and F. David Schoorman (1995), "An Integrative Model of Organizational Trust," *Academy of Management Review*, 20(3), 709–34.

McFarland, Matt (2018), "Uber Self-Driving Car Kills Pedestrian in First Fatal Autonomous Crash," *CNN*, https://money.cnn.com/2018/03/19/technology/uber-autonomous-car-fatal-crash/index.html.

McKnight, D. Harrison and Norman L. Chervany (2001), "Trust and Distrust Definitions: One Bite at a Time," in *Trust in Cyber-Societies*, ed. Rino Falcone, Munindar Singh, and Yao-Hua Tan, Berlin, Heidelberg: Springer, 27–54.

Meehl, Paul E. (1954), *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.*, University of Minnesota Press, University of Minnesota Press.

Moray, Neville, Toshiyuki Inagaki, and Makoto Itoh (2000), "Adaptive Automation, Trust, and Self-Confidence in Fault Management of Time-Critical Tasks," *Journal of Experimental Psychology: Applied*, 6(1), 44–58.

Muir, Bonnie M. (1987), “Trust between Humans and Machines, and the Design of Decision Aids,” *International Journal of Man-Machine Studies*, 27(5–6), 527–39.

Muir, Bonnie M and Neville Moray (1996), “Trust in Automation. Part II. Experimental Studies of Trust and Human Intervention in a Process Control Simulation,” *Ergonomics*, 39(3), 429–60.

Newman, David T., Nathanael J. Fast, and Derek J. Harmon (2020), “When Eliminating Bias Isn’t Fair: Algorithmic Reductionism and Procedural Justice in Human Resource Decisions,” *Organizational Behavior and Human Decision Processes*, 160(April 2018), 149–67.

Nisbett, Richard E. and Timothy Decamp Wilson (1977), “Telling More than We Can Know: Verbal Reports on Mental Processes,” *Psychological Review*, 84(3), 231–59.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan (2019), “Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations,” *Science*, 366(6464), 447–53.

Önkal, Dilek, M. Sinan Gönül, and Shari De Baets (2019), “Trusting Forecasts,” *Futures & Foresight Science*, 1–10.

Önkal, Dilek, Paul Goodwin, Mary Thomson, Sinan Gönül, and Andrew Pollock (2009), “The Relative Influence of Advice from Human Experts and Statistical Methods on Forecast Adjustments,” *Journal of Behavioral Decision Making*, 22(4), 390–409.

Poursabzi-Sangdeh, Forough, Daniel G. Goldstein, Jake M. Hofman, Jennifer Wortman Vaughan, and Hanna Wallach (2018), “Manipulating and Measuring Model Interpretability,” *arXiv preprint*.

Prahl, Andrew and Lyn Van Swol (2017), “Understanding Algorithm Aversion: When Is Advice from Automation Discounted?," *Journal of Forecasting*, 36(6), 691–702.

Promberger, Marianne and Jonathan Baron (2006), "Do Patients Trust Computers?," *Journal of Behavioral Decision Making*, 19(5), 455–68.

Rice, Lisa and Deidre Swesnik (2013), "Discriminatory Effects of Credit Scoring on Communities of Color," *Suffolk University Law Review*, 46(3), 935–65.

Rousseau, Denise M, Sim B Sitkin, Ronald S Burt, and Colin Camerer (1998), "Not so Different after All: A Cross-Discipline View of Trust," *Academy of Management Review*, 23(3), 393–404.

Rudin, Cynthia (2019), "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence*, 1(5), 206–15.

Ryan, Richard M. and Edward L. Deci (2000), "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being.," *The American Psychologist*, 55(1), 68–78.

Sapienza, Harry J . and Audrey Korsgaard (1996), "Procedural Justice in Entrepreneur-Investor Relations," *The Academy of Management Journal*, 39(3), 544–74.

Schmitt, Bernd (2019), "From Atoms to Bits and Back: A Research Curation on Digital Technology and Agenda for Future Research," *Journal of Consumer Research*, 46(4), 825–32.

Shariff, Azim, Jean François Bonnefon, and Iyad Rahwan (2017), "Psychological Roadblocks to the Adoption of Self-Driving Vehicles," *Nature Human Behaviour*, 694–96.

Sloman, Steven A. and Nathaniel Rabb (2016), "Your Understanding Is My Understanding: Evidence for a Community of Knowledge," *Psychological Science*, 27(11), 1451–60.

Spiller, Stephen A. and Lena Belogolova (2017), "On Consumer Beliefs about Quality and Taste," *Journal of Consumer Research*, 43(6), 970–91.

Supreme Court of Wisconsin (2016), "State of Wisconsin v. Eric L. Loomis," https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690.

The Economist (2016), "Streaming on Screens near You," *The Economist*, August 20, 1–5.

The United States Congress (2011), *15 U.S.C.-Equal Credit Opportunity Act*, The U.S. Government Publishing Office.

U.S. Department of Transportation's National Highway Traffic Safety Administration (2020), *Fatal Motor Vehicle Crashes in 2019*.

Wakabayashi, Daisuke (2018), "Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam," *New York Times*.

Wilson, Timothy Decamp (2004), *Strangers to Ourselves: Discovering the Adaptive Unconscious*, Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Yaniv, Ilan (2004), "Receiving Other People's Advice: Influence and Benefit," *Organizational Behavior and Human Decision Processes*, 93(1), 1–13.

Yaniv, Ilan and Eli Kleinberger (2000), "Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation," *Organizational Behavior and Human Decision Processes*, 83(2), 260–81.

Yeomans, Michael, Anuj Shah, Sendhil Mullainathan, and Jon Kleinberg (2019), "Making Sense of Recommendations," *Journal of Behavioral Decision Making*, 32(4), 403–14.

Yu, Martin C. and Nathan R. Kuncel (2020), "Pushing the Limits for Judgmental Consistency: Comparing Random Weighting Schemes with Expert Judgments," *Personnel Assessment and Decisions*, 6(2).