

When Algorithms Fail

Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors

Srinivasan, Raji ; Abi, Gülen Sarial

Document Version

Accepted author manuscript

Published in:

Journal of Marketing

DOI:

[10.1177/0022242921997082](https://doi.org/10.1177/0022242921997082)

Publication date:

2021

License

Unspecified

Citation for published version (APA):

Srinivasan, R., & Abi, G. S. (2021). When Algorithms Fail: Consumers' Responses to Brand Harm Crises Caused by Algorithm Errors. *Journal of Marketing*, 85(5), 74-91. <https://doi.org/10.1177/0022242921997082>

[Link to publication in CBS Research Portal](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us (research.lib@cbs.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. Jul. 2025



Author Accepted Manuscript



**When Algorithms Fail: Consumers' Responses to Brand
Harm Crises Caused by Algorithm Errors**

Journal:	<i>Journal of Marketing</i>
Manuscript ID	JM.20.0236.R4
Manuscript Type:	Revised Submission
Research Topics:	Brand Management, Crisis Management
Methods:	Lab Experiments

SCHOLARONE™
Manuscripts

**When Algorithms Fail: Consumers’ Responses to Brand Harm Crises Caused by
Algorithm Errors**

Abstract

Algorithms increasingly used by brands sometimes fail to perform as expected or even worse, cause harm, causing brand harm crises. Unfortunately, algorithm failures are increasing in frequency. Yet, we know little about consumers’ responses to brands following such brand harm crises. Extending developments in the theory of mind perception, we hypothesize that following a brand harm crisis caused by an algorithm error (vs. human error), consumers will respond less negatively to the brand. We further hypothesize that consumers’ lower mind perception of agency of the algorithm (vs. human) for the error that lowers their perceptions of the algorithm’s responsibility for the harm caused by the error will mediate this relationship. We also hypothesize four moderators of this relationship: two algorithm characteristics, anthropomorphized algorithm and machine learning algorithm and two task characteristics where the algorithm is deployed, subjective (vs. objective) task and interactive (vs. non-interactive) task. We find support for the hypotheses in eight experimental studies including two incentive-compatible studies. We examine the effects of two managerial interventions to manage the aftermath of brand harm crises caused by algorithm errors. The research’s findings advance the literature on brand harm crises, algorithm usage, and algorithmic marketing and generate managerial guidelines to address the aftermath of such brand harm crises.

Keywords: brand harm crises, algorithmic marketing, algorithm errors, theory of mind perception

Author Accepted Manuscript

Given the explosive growth in the volume of data, dramatic developments in software programs, and the decreasing cost of cloud computing, the usage of algorithms, software programs that organize data, predictions, and decisions, has grown exponentially. While this has occurred across many contexts, algorithm usage in the marketing context, algorithmic marketing has increased dramatically. Algorithmic marketing has many advantages including lower costs, high efficiency, and effectiveness (Gal and Elkin-Koren 2017). Despite their advantages, there is growing evidence of algorithm failures across multiple contexts (Griffith 2017). In the marketing context, algorithm errors harm consumers and/or violate consumers' expectations of the brand's values, creating brand harm crises. In a survey of Chief Marketing Officers (CMOs), fielded by the CMO Council and Dow Jones Inc. (2017), most CMOs (78%) expressed concern about the threats to their brands' reputations from algorithm errors.

Although algorithms operate in the digital domain, algorithm errors have many real-world consequences, including causing substantive harm to brands. We discuss two examples to provide additional context. First, there is evidence (Diakopoulos 2013; Sweeney 2013) of algorithmic defamation in online searches. Algorithm-based Google search auto-completion routines make incorrect defamatory associations about groups of people (Badger 2019). For example, searching for certain ethnic names on Google provides results of advertising for bail bonds or criminal record checking. Second, Apple Credit Card, launched in partnership by Apple Inc. and Goldman Sachs Inc. in August 2019, faced reputational harm when users noticed that it offered lower lines of credit to women than to men of equal or even lower financial standing (Vigdor 2019). In response, the New York Department of Financial Services announced an investigation of Apple Inc. to assess a breach of federal financial rules on equal financial access. Cognizant of the potential harm from algorithm errors, for the first time, Google's

parent company, Alphabet Inc. (February 2019) and Microsoft Inc. (August 2018) acknowledged in their annual reports that “flawed” algorithms could result in “brand or reputational harm” and have an “adverse affect” on financial performance (Vincent 2019). In sum, algorithm errors are a key and growing source of brand harm crises.

Brand harm crises are adverse negative events inconsistent with a brand’s values. In a brand harm crisis, the brand’s ability to deliver promised benefits to consumers is compromised or even worse causes physical harm to consumers (Dutta and Pullig 2011; Pullig, Netemeyer, and Biswas 2006) so that consumers respond negatively to the brand (Ahluwalia, Burnkrant, and Unnava 2000; Lei, Dawar, and Gürhan-Canli 2012; Swaminathan, Page, and Gürhan-Canli 2007). Consumers’ attributions about what caused the harm influence their subsequent responses to the brand (Folkes 1984; 1990). Consumers feel angry and seek revenge if they believe that the firm was responsible for the harm and could have prevented it (Folkes, Koletsky, and Graham 1987). See Cleeren, Dekimpe, and Heerde (2017) for a comprehensive review of the brand harm crises literature. Given the recent growth in algorithmic marketing, extant research has overlooked harm crises caused by algorithm errors.

There is a large body of research in multiple literatures, including in marketing, on people’s responses to nonhuman agents (e.g., algorithms, computers, robots, etc.). People treat computers as social actors although they know that computers do not possess feelings, intentions, motivations or “selves” (Moon 2000; Nass and Moon 2000). Other work (Choi, Matilla and Bolton, forthcoming) suggests that humanoid (vs. non-humanoid) service robots are more strongly associated with warmth (whereas competence is not).

Past work on algorithm usage has examined people’s responses to using algorithms (Castelo, Bos, and Lehmann 2019; Dietvorst, Simmons, and Massey 2015). Individuals prefer

doing a task themselves or having it done by their peers (than by algorithms) with whom they have more in common (Prahl and van Swol 2017) than using an imperfect algorithm, i.e., people display algorithm aversion (Dietvorst, Simmons, and Massey 2015). This preference for using humans over algorithms persists even when doing so, worsens outcomes. In contrast, in the advice-giving context (absent of algorithm errors), Logg, Minson, and Moore (2019) report algorithm appreciation, i.e., people incorporate advice from algorithms more than from humans. Related recent work on automated vehicles operated by algorithms (Awad et al. 2020; Gill 2020) suggests that individuals considered harm to pedestrians by an automated vehicle (vs. themselves as the driver in a regular car) more permissible. Please see Castelo, Bos, and Lehmann (2019) for a good overview of the research on algorithm usage (Table 1 on p. 2).

In sum, past research on algorithms has overlooked how consumers will respond to a brand following a brand harm crisis caused by an algorithm error (vs. human error), the focus of this research. Distinct from past research on algorithm usage which considers the individual's decision to use the algorithm, in this research context, the decision to use the algorithm is taken by the brand manager not by the consumer who experiences the harm caused by the algorithm error. Further, the dependent variable here is the consumer's response to the brand and not to the algorithm that commits the error, the focus of past research on algorithm usage. Further, we examine the moderation effects of two algorithm characteristics and two task characteristics where the error occurs on this relationship. As consumers' responses to a brand harm crisis are always negative (Lei, Dawar, and Gürhan-Canli 2012), we examine consumers' negative responses to a brand following a brand harm crisis caused by an algorithm error.

We apply the theory of mind perception (Gray, Gray, and Wegner 2007; Gray and Wegner 2012) that individuals ascribe minds to other entities (e.g., individuals, animals, and

robots) and reason about the contents of these entities’ minds. Specifically, we consider consumers’ mind perception of agency of the algorithm, i.e., the entity’s perceived capacity to intend and to act that has committed an error.

Features of an entity can change people’s mind perception of the entity’s agency (Waytz, Cacioppo, and Epley 2010). Accordingly, we hypothesize that following a brand harm crisis caused by an algorithm error (vs. human error), consumers will, *ceteris paribus*, have lower mind perception of agency of the algorithm (the entity) and assign it lower responsibility for the harm caused, weakening their negative responses to the brand. Further, individuals’ responses to an algorithm vary based on the task characteristics (Castello et al. 2019). Accordingly, we consider four moderators of consumers’ responses to a brand following a harm crisis caused by an algorithm error: two algorithm characteristics, anthropomorphized algorithm and machine learning algorithm and two task characteristics where the algorithm error occurs, subjective (vs. objective) task and interactive (vs. non-interactive) task. We test and find support for the hypotheses in eight experimental studies, including an incentive compatible study with a consequential outcome (donation to a charity) and two studies with behavioral measures.

This research’s insights extend the literature on harm crises by studying an inanimate source of errors, algorithms, hitherto overlooked in the marketing literature. Second, in a novel extension to the algorithm usage literature which has hitherto focused on consumers’ responses to the algorithm, consumers responses to the brand are more forgiving of algorithm errors when they do not have the authority on whether to use the algorithm or not. Third, we identify consumers’ mind perception of agency of algorithms as a potential key building block, relevant in the development of a theory of algorithmic marketing. Fourth, by identifying the moderating role of algorithm and task characteristics, this research’s insights make a novel contribution to

the harm crises literature that has not examined characteristics of the sources of the error and task as factors affecting outcomes in harm crises. Using the insights from the findings of the four moderators and a managerial interventions study, we provide guidance to managers on the deployment of algorithms, given their effects on consumers' responses when they commit errors, and how to manage the aftermath of such brand harm crises.

Theory

Early work on people's responses to nonhuman agents (e.g., computers) suggests that consumers mindlessly apply social norms (Moon 2000) in their interactions with computers including displaying a self-serving bias in attributions of responsibility to positive versus negative service encounters (Moon 2003). Building on these ideas, we apply the theory of mind perception in the psychology literature (Gray, Gray, and Wegner 2007; Gray and Wegner 2012) about people's perceptions of the minds of other entities to algorithms to develop the hypotheses. We first provide a brief overview of the theory of mind perception and then develop the hypotheses.

Theory of Mind Perception: A Brief Overview

Mind perception, also known as humanizing or mentalizing, involves making inferences about one's own and others' (entities) mental states by positing unobservable properties such as intentions, desires, goals, beliefs, and secondary emotions to serve as mediators between people's sensory inputs and their subsequent actions (Gray, Gray, and Wegner 2007; Gray and Wegner 2012). According to the theory of mind perception, a perceiver needs to implicitly determine the extent to which an entity has a mind and then determine that entity's state of mind. In addition, to perceiving the minds of other humans, people are capable of perceiving minds of non-human entities such as animals, gadgets, or software.

People represent other entities’ minds on two psychological capacities, agency and experience (Gray, Gray, and Wegner 2007). Mind perception of the entity’s agency is its perceived capacity to intend and to act (e.g., self-control, judgment, communication, thought, and memory) and mind perception of experience is the entity’s perceived capacity for sensation and feeling (e.g., hunger, fear, pain, pleasure, and consciousness) that are acted upon the entity. When discussing mind perception of agency, Gray, Gray, and Wegner (2007) posit that agency qualifies entities as moral agents, capable of reasoned actions and with the capacity to do right or wrong (Gray and Wegner 2009; Gray, Young, and Waytz 2012) whereas experience qualifies entities as moral patients, capable of benefiting from good or suffering from evil acted upon them.¹

In this research, we consider consumers’ mind perception of agency of the algorithm that has committed the error and do not consider consumers’ mind perception of the algorithm’s experience, as a moral patient, being acted upon by others, which is not relevant when the algorithm commits errors. We note that individuals’ mind perception of agency of an entity are positively related to judgments of the entity’s responsibility for harm caused (Waytz, Heafner, and Epley 2014), which is consistent with common law practice that holds individuals with diminished mental capacity as being less responsible for their transgressions.

Overview of Hypotheses

We propose that following a brand harm crisis caused by an algorithm error (vs. human error), consumers will have lower mind perception of agency of the algorithm (than humans) for

¹ The common everyday meaning of “experience” as “practical contact with and observation of facts or events” (from the Merriam Webster dictionary) is distinct the use of the term “experience” in the theory of mind perceptions, defined as the capacity for sensations (i.e., felt by the algorithms).

the error, assign lower responsibility to the algorithm for the harm caused by the error, resulting in a less negative response to the brand.

Features of the entity can change people's mind perception of its agency (Waytz, Cacioppo, and Epley 2010). Further, individuals' responses to algorithms varies based on the characteristics of the task for which the algorithm is deployed (Logg, Minson, and Moore 2019). Extending these two ideas, we propose four factors that will moderate consumers' responses to a brand following a brand harm crisis caused by an algorithm error: two algorithm characteristics, anthropomorphized algorithm and machine learning algorithm and two task characteristics where the error occurs, subjective (vs. objective) task and interactive (vs. non-interactive) task.

Main Effect of Algorithm (vs. Human) Error

An entity's mind perception of agency to intend and to act affect individuals' perception of the entity's responsibility for its actions. For example, people have lower mind perception of agency of an inanimate robot than of a man or of a young girl (Gray, Gray, and Wegner 2007), suggesting perception of lower responsibility for the robot's harmful actions. Extending this idea to algorithm errors, we propose that people will have lower mind perception of the agency of the algorithm (vs. human) which commits the error that causes the brand harm crisis and assign lower responsibility² to the algorithm for the harm caused.

²We note that the meaning of the term "responsibility" has three commonplace meanings (Mirriam Webster dictionary): 1) the state or fact of having a duty to deal with something or of having control over someone, 2) the state or fact of being accountable or to blame for something, and 3) the opportunity or ability to act independently and make decisions without authorization. Our usage of the term "responsibility" is as per the definition in point 2 above. Consistent with this interpretation, "blame" is a synonym for "responsibility" at <https://www.merriam-webster.com/dictionary/responsibility>. Thus, our view of responsibility for the harm is consistent with blame for the harm.

That people consider algorithms to have lower agency than humans who developed the algorithm is consistent with early research on individuals’ interactions with computers and robots (Moon 2000) and the recent research on algorithm aversion (Dietvorst et al. 2015) and algorithm appreciation (Logg et al. 2019; Prahl and van Swol 2017). This argument is consistent with other evidence on algorithms (McCullom 2017) that as algorithms do not have “human-like” qualities; people may not hold them fully responsible for actions that cause harm.

Accordingly, we propose that consumers’ responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) will be less negative. We further propose that consumers’ responses to the brand will be serially mediated by their lower mind perception of the algorithm’s agency which, in turn, will lower their perceptions of the algorithm’s responsibility for the harm caused by the error. Hence, we propose H₁ and H₂:

H₁: Consumers’ responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) will be less negative.

H₂: Consumers’ lower mind perception of the algorithm’s agency, which will lower their perceptions of the algorithm’s responsibility for the harm caused by the error, will mediate the relationship in H₁.

Anthropomorphized Algorithm

Anthropomorphism is the process of inductive inference where people attribute distinctively human characteristics to inanimate objects, including brands, machines, technologies, and software (Kim and McGill 2011). Anthropomorphizing an entity includes the use of human characteristics (e.g., human-like face and name) so that individuals attribute essential human characteristics (e.g., human-like mind capable of thinking and feeling) to the entity. A common marketing practice is to name products with human names with the intent of

anthropomorphization (e.g., IBM's artificial intelligence software "Watson," Bank of America's virtual financial assistant "Erica", and Amazon's virtual assistant "Alexa").

The effects of anthropomorphization on consumer behaviors have received attention from marketing scholars (Aggarwal and McGill 2007; 2012; Kim and Kramer 2015). The overall evidence suggests that the higher a product's anthropomorphization, the higher consumers' evaluations of it (Aggarwal and McGill 2007) and the higher its sales (Landwehr, McGill, and Herrmann 2011). With regard to harm crises, anthropomorphization of a product that humanizes it lowers its consumers' evaluations (Puzakova, Kwak, and Rocereto 2013) which is consistent with the main effect (H_1) above.

In the technology context, relevant to this research, firms anthropomorphize products to make them user friendly and less intimidating (Lafrance 2014). Anthropomorphizing technology-driven products increases consumers' positive feelings toward the products, reduces people's fear of technology, suggests that the products can perform their intended functions well (Waytz, Heafner, and Epley 2014). This results in assigning higher responsibility to anthropomorphized products, indeed, at a level comparable to those of humans (Epley, Caruso, and Bazerman 2006).

Accordingly, we suggest that when an anthropomorphized (vs. not) algorithm is the source of the error that causes a brand harm crisis, consumers will consider the anthropomorphized algorithm to have higher mind perception of agency and assign higher responsibility to it for the harm caused by the algorithm error. We hypothesize that consumers' responses to a brand following a brand harm crisis caused by an algorithm error will be more negative when the algorithm is anthropomorphized (vs. not). Hence, we propose H_3 :

H₃: Consumers’ responses to a brand following a brand harm crisis will be more negative when the error is caused by an anthropomorphized (vs. not) algorithm.

Machine Learning Algorithm

Machine learning algorithms learn “by themselves” i.e., independently, using historical data, models, and analyses. In other words, the machine learning algorithm is programmed such that it can modify itself (i.e., without human intervention) to improve its performance. The availability of ‘Big Data,’ growing computational power, and developments in software technology, enable such machine learning algorithms to learn independently from their experiences working repeatedly on large datasets (Heller 2019). Machine learning algorithms know users’ behaviors and leverage that knowledge to recommend products that match users’ preferences. Such machine learning algorithms power Amazon, Netflix, and Spotify recommendations, Google Maps, and much of the content on Facebook, Instagram, and Twitter.

Developments in bio-ethics consider an entity’s capacity for learning, including the ability to think, to reason, and remember as having superior mental abilities and defining the degree of its humanness (Fletcher 1979). Reiterating this view, Gray and Wegner (2009) compared people’s perceptions of mentally competent (vs. mentally challenged) adults and found them to be higher on mental abilities associated with learning and mind perception of agency.

Applying these ideas, we propose that consumers will ascribe more humanness to a machine learning (vs. not) algorithm. Following a brand harm crisis caused by an error of a machine learning (vs. not) algorithm, people may perceive the machine learning algorithm to have higher agency and therefore, higher responsibility for the harm caused. Thus, we hypothesize that following a brand harm crisis caused by an error of a machine learning (vs. not) algorithm, consumers’ responses to the brand will be more negative. Hence, we propose H₄:

H₄: Consumers' responses to a brand following a brand harm crisis will be more negative when the error is caused by a machine learning (vs. not) algorithm.

Subjective (vs. Objective) Task

Following Castelo, Bos, and Lehmann (2019), a subjective task is open to interpretation based on an individual's personal opinion while an objective task is one that involves factors that are quantifiable and measurable. People perceive subjective tasks as requiring intuition and objective tasks as requiring human traits as logical, rule-based analysis (Inbar, Cone, and Gilovich 2010). Although algorithms are proficient at objective tasks, the growth of 'Big Data' and lower costs of computing has resulted in a dramatic increase in the use of algorithms for subjective tasks (Kleinberg et al. 2018). Companies routinely use algorithms for subjective tasks, such as selecting applicants (e.g., Indeed.com, University admissions) and personal wardrobes for consumers (e.g., J. Jills, Stitchfix.com). Ceteris paribus, consumers perceive that algorithms lack abilities to perform subjective tasks (Castelo, Bos, and Lehmann 2019) although increasing the algorithm's human-likeness is effective at increasing its usage for subjective tasks. In other words, individuals ascribe higher humanness to an algorithm deployed for subjective tasks.

Applying the above logic, we propose that when the algorithm is used in a subjective (vs. objective) task which requires intuition and an algorithm error causes the brand harm crisis, consumers will perceive the algorithm as having higher mind perception of agency and hold it more responsible for the harm caused. Thus, we propose that in a brand harm crisis caused by an algorithm error, when the algorithm error occurs in a subjective (vs. objective) task, consumers' responses to the brand will be more negative. Hence, we propose H₅:

H₅: Consumers' responses to a brand following a brand harm crisis will be more negative when the algorithm error occurs in a subjective (vs. objective) task.

Interactive (vs. Non-Interactive) Task

A key characteristic of interactive communications between entities (say, a human and a computer) is contingency in responses (Sundar 2009). In an interactive communication (Sundar et al. 2016), each entity acknowledges and incorporates the other entity’s prior communications. Higher interactivity between two individuals in an online context heightens perceptions of each other’s humanness (Sundar et al. 2015). Interactivity between an individual and a non-human entity (e.g., an algorithm) makes the entity more human because it mimics the contingency in real-time interactive exchanges between humans (Rafaeli 1988). Hence, people may perceive the algorithm in an interactive task as being capable of communication, an integral aspect of people’s mind perception of agency of an entity (Gray, Young, and Waytz 2012). Indeed, algorithms are now widely used by marketers in interactive communications including in customer service chatbots (e.g., Spotify) and product recommendations (e.g., Stitchfix).

Applying these ideas, we anticipate that consumers will have higher mind perception of agency in an interactive (vs. non-interactive) task between consumers and the algorithm in the task where the algorithm error occurs. We propose that, following a brand harm crisis caused by an algorithm error in an interactive (vs. non-interactive) task, consumers will hold the algorithm more responsible for the harm caused, so that consumers’ responses to the brand, following the brand harm crisis, will be more negative. Hence, we propose H₆:

H₆: Consumers’ responses to a brand following a brand harm crisis caused by an algorithm error will be more negative when the error occurs in an interactive (vs. non-interactive) task.

Pre-study

We conducted a pre-study that examines consumers' responses to a brand when there are no errors to ensure that the effects, that we theorize above, relate only to the error caused by the algorithm (vs. human) and not more generally to algorithms.³ We pre-registered this pre-study on AsPredicted.org (#43436). We provide stimuli for the pre-study and all other studies in the Web Appendix and summary of the studies and findings in Table 1.

---- Insert Table 1 ----

Participants and Procedure

Four hundred and three adults participated in the experiment on MTurk in exchange for 50 cents (219 male; $M_{\text{age}} = 37.73$, $SD = 12.40$). The study used error (vs. no error) and algorithm (vs. human) between-subjects design.

We randomly assign participants to error (vs. no error) conditions. Participants in the error condition read that HMS Investments, a leading financial investment company, was facing a crisis. In the 'no error' condition, participants read that HMS Investments reduces risks for its

³ The Institutional Review Board of the authors' home institutions reviewed and approved the experimental design before commencing the research. Participants in all studies provided informed consent before participation.

As an empirical practice, we had a rule of thumb of ensuring at least 30 participants per cell for lab studies and at least 75 participants per cell for online studies. For the lab studies using student participants, we did not know, a priori, the number of participants. We report all variables collected and all conditions in the studies and do not exclude data from the analyses unless otherwise noted for clearly identified reasons. We report the number of excluded participants and do not add data from additional participants in any study, following the analyses.

We pre-registered analyses (and exclusions) at AsPredicted.org for Pre-study, Studies 3, 5 and the managerial intervention study. We conducted Studies 1a-1c, 2, 4, and 6 before pre-registration became our standard practice. Anonymized links to the preregistrations of studies are available upon request from the authors.

We conducted all analyses on SPSS Statistics 23 and 25 IBM software.

clients. We randomly assign participants to algorithm (vs. human) conditions. Participants in the algorithm (vs. human) error condition read that HMS Investments, a leading financial investment company, was facing a crisis because a financial algorithm program (financial manager) had committed an error, resulting in financial losses for its customers. Participants in the algorithm (vs. human) ‘no error’ condition read that HMS Investments reduces risks for its clients with its strong computer algorithms (vs. employees). We measured participants’ attitude toward the HMS Investments using a five-item scale (Ahluwalia, Burnkrant, and Unnava 2000; Swaminathan, Page, and Gürhan-Canli 2007): bad/good, low quality/high quality, undesirable/desirable, harmful/beneficial, unfavorable/favorable ($\alpha = .96$). Participants then provided their basic demographic information.

Results

Brand attitude. An ANOVA analysis on brand attitude reveals the predicted interaction effect of error (vs. no error) and algorithm (vs. human) conditions, $F(1, 399) = 5.86, p = .016$. There is no main effect of error (vs. no error), $p = .157$, and algorithm (vs. human) conditions, $p = .335$. Participants’ responses to a brand following a brand harm crisis caused by an algorithm (vs. human) error are less negative, $M_{AE} = 4.55, SD = 1.56$ vs. $M_{HE} = 3.63, SD = 1.79, F(1,399) = 21.63, p < .001$. However, participants’ responses to a brand that uses algorithms (vs. humans) are not different when there is no error, $M_{ALGORITHM} = 5.55, SD = 1.03$ vs. $M_{HUMAN} = 5.31, SD = 1.07, F(1,399) = 1.53, p = .217$. There is no effect of age, $p = .085$, or gender, $p = .612$.

Thus, consumers’ responses to a brand that uses an algorithm (vs. human) when there is no error are not different. However, as hypothesized in H_1 , consumers’ responses to a brand following a brand harm crisis caused by an algorithm (vs. human) error are less negative. These

findings provide preliminary support for H_1 and indicate that algorithm error, and not the mere presence of the algorithm, drives the results in the subsequent studies.

Study 1a-1c: Main Effect of Algorithm Error (vs. Human Error)

Study 1a

In study 1a, we examine consumers' responses to a brand following brand harm crisis caused by an algorithm error (vs. human error) with an incentive compatible experiment using a consequential outcome (donation to a charity suggested by the brand), as the dependent variable. A lower donation to the charity denotes a more negative response to the brand.

Participants read about a consumer electronics retailer where an algorithm error or a human error had caused a brand harm crisis. Participants then indicated the amount that they were willing to donate to the World Health Organization, through the electronics retailer, from the compensation that they would receive in the study.

Participants and Procedure

One hundred and fifty-seven US adults participated in the experiment on MTurk in exchange for 150 cents (84 male; $M_{age} = 40.69$, $SD = 10.37$). All participants read that a consumer electronics company, Qualtronics, was facing a harm crisis. This was because their fund-raising campaign, BanishCovid19, aimed at combatting Covid 19, implied that a Chinese virus caused Covid 19. The fund-raising campaign was for the World Health Organization.

Participants in the algorithm error condition read: Because the disease was first detected in Wuhan Province of China, Qualtronics used computer algorithms to design the advertisement and released the campaign with the headline "Contribute to BanishCovid19 and Destroy the Chinese Virus." Participants in the human error condition read: Because the disease was first detected in Wuhan Province of China, Qualtronics' managers designed the advertisement and

released the campaign with the headline “Contribute to BanishCovid19 and Destroy the Chinese Virus.” In both conditions, participants read that following negative feedback from their customers, Qualtronic apologized to its customers and changed the advertisement headline to “Contribute to BanishCovid19 and Destroy the Corona Virus.”

We then provided participants in both conditions, the opportunity to donate to the World Health Organization through Qualtronic. The maximum amount that they could donate was the 150 cents that they would earn in the study. Participants indicated the amount that they would donate on a sliding scale ($M = 14.40$, $SD = 34.47$).

As a manipulation check, participants indicated the extent to which they believed that the error was caused by a human in Qualtronic (1 = not at all and 7 = very much). Participants also indicated the extent to which they were concerned about COVID 19 and the extent to which COVID 19 had impacted their community on seven-point scales (1 = not at all and 7 = very much). Participants then provided their basic demographic information including race.

Results

Manipulation check. As intended, participants in the human error (vs. algorithm error) condition indicated that the source of the error in Qualtronic is more human, $M_{HE} = 6.29$, $SD = 1.34$ vs. $M_{AE} = 5.81$, $SD = 1.57$, $t(155) = 2.03$, $p = .044$.

Amount of donation. The results indicate a significant effect of algorithm error (vs. human error) condition on the donation amount, $M_{AE} = 20.71$, $SD = 41.91$ vs. $M_{HE} = 7.84$, $SD = 22.34$, $F(1, 155) = 5.70$, $p = .018$. When we included the three control variables of participants’ race, concerns about COVID 19, or COVID 19’s impact on their community as control variables in the model, the effect of algorithm error (vs. human error) on the amount of donation is still

significant, $F(1, 152) = 5.11, p = .024$. There was no main effect of the three control variables or of age, $p = .289$, and gender, $p = .202$.

In Study 1a, consumers' donation of money following a brand harm crisis caused by an algorithm error (vs. human error) are higher. This finding supports the prediction (H_1) that consumers' responses to the brand following a brand harm crisis caused by an algorithm error (vs. human error) are less negative.

Study 1b

In study 1b, we examine consumers' behavioral responses to a brand harm crisis caused by an algorithm error (vs. human error) at a fictitious global platform company, Life Skills without Borders, an online advice crowdsourcing website for young adults. We randomly assigned participants to either the algorithm error or human error condition and measured the number of items of advice provided by participants to Life Skills without Borders, following a brand harm crisis.

Participants and Procedure

The experiment used the algorithm error (vs. human error) condition as a between-subjects design. Two hundred and thirty-three participants participated in the experiment on the Prolific online platform in exchange for one British pound (101 male; $M_{\text{age}} = 35.89$, $SD = 12.36$).

All participants read about Life Skills without Borders, a global crowdsourcing platform for providing life skills advice to young adults. We randomly assigned participants to either the algorithm error or human error conditions. Participants in the algorithm error (human error) condition read that a computer algorithm (an employee) at Life Skills without Borders had made

Author Accepted Manuscript

a mistake and provided wrong financial advice to poor young couples, resulting in financial losses.

We then informed participants that Life Skills without Borders was presently crowdsourcing ideas for providing post-graduation career advice to young adults. We asked participants to provide advice to Life Skills without Borders, which was the study’s dependent variable. We used the number of unique items of advice provided by each participant (e.g., (1) Follow your heart and work at a job that you think you might like, (2) If you get an offer for a higher paying job at a different company be sure you want the job before you take it) as the dependent variable. As the only difference between the two (between subject) conditions was the source of the error (algorithm vs. human), we consider the higher number of pieces of advice provided by participants as indicative of participants’ less negative response to Life Skills. Participants then provided their basic demographic information.

Results and Discussion

Career Advice. One of the authors coded the number of unique items of advice provided by the participants ($M_{ADVICE\#} = 2.52$, $SD = 2.13$). A t-test shows that participants in the algorithm error condition provided more advice than did participants in the human error condition, $M_{AE} = 3.19$, $SD = 2.18$ vs. $M_{HE} = 2.49$, $SD = 1.91$, $t(229) = 72.56$, $p = .011$.

The results of study 1b support our prediction (H_1) that consumers’ behavioral responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) are less negative.

Study 1c

Author Accepted Manuscript

In study 1c (details of study in the Web Appendix), we examine consumers' re-engagement behaviors with the brand following a brand harm crisis caused by a failure in the online computer system. An algorithm error (vs. human error) disrupted the online task on Qualtrics, the software program used for lab experiments. We randomly assigned participants either to the algorithm error or human error condition and noted their decision to repeat or not repeat the online task (i.e. re-engage with Qualtrics). Participants' willingness to repeat the task would indicate a less negative response to the Qualtrics brand. The results support our prediction (H_1) that consumers' re-engagement behaviors with the brand (i.e., repeat the online task), following a brand harm crisis caused by an algorithm error (vs. human error) are less negative.

Study 2: Mediation by Mind Perception of Algorithm's Agency and Responsibility for Harm

In study 2, we examine the role of consumers' mind perception of the source of the error's agency in committing the error and responsibility for the harm caused in serially mediating consumers' responses to the brand following a brand harm crisis caused by an algorithm error (vs. human error) (H_2).⁴ For a test of the mediation (H_2), we measured participants' mind perception of agency of the source of the error that caused the brand harm crisis and perceptions of the source of the error's responsibility for the harm caused by the error.

Participants and Procedure

Two hundred and fifty-one adults participated in the between-subjects experiment on MTurk online platform in exchange for 50 cents (137 male; $M_{age} = 34.98$, $SD = 11.19$). All

⁴ Following the suggestion of a reviewer, we conducted a pre-test which rules out the alternative explanation that people attribute more agency to algorithms so that when algorithms make mistakes they may consider that "even a superior entity that has higher capacities made a mistake". Ruling out this explanation, a pretest ($N=153$) indicated that people significantly attribute more agency to humans than they do to algorithms ($M_{HUMAN} = 5.95$, $SD = .94$, $M_{ALGORITHM} = 4.47$, $SD = 1.42$, $t(152) = 10.409$, $p < .001$).

participants saw a tweet on the official Twitter account of the New York Times website announcing the recall of 4.8 million Fiat Chrysler vehicles because of a cruise control problem. We randomly assigned participants to either the algorithm error or human error condition. Participants in the algorithm error (human error) condition read that a computer algorithm (Fiat Chrysler employees) at Fiat Chrysler had made a mistake resulting in a defect in the cruise control system causing a safety hazard.

We measured participants' attitude toward the Fiat Chrysler brand using a five-item scale (Ahluwalia, Burnkrant, and Unnava 2000; Swaminathan, Page, and Gürhan-Canli 2007): bad/good, low quality/high quality, undesirable/desirable, harmful/beneficial, unfavorable/favorable ($\alpha = .96$). We measured participants' mind perception of agency of the source of error using Gray, Gray, and Wegner's (2007) seven-item scale. The items are: 1) telling right from wrong 2) remembering things 3) understanding how others feel 4) conveying thoughts to others 5) of making plans 6) exercising self-restraint over impulses, and 7) thinking (1 = not at all and 7 = very much; $\alpha = .95$). We then measured participants' perceptions of the source of the error's responsibility for the harm caused by the error using Waytz, Heafner, and Epley's (2014) four-item scale. The items are the extent to which the source of the error at Fiat Chrysler 1) was responsible 2) must be held to account 3) deserves blame and 4) was blameworthy for the harm caused by the error (1 = not at all and 7 = very much; $\alpha = .93$).

As a manipulation check, we asked participants to indicate the extent to which they thought that the source of the error was a human and the extent to which they thought that the source of the error was a computer algorithm (1 = not at all and 7 = very much). Finally, participants provided their basic demographic information.

Results

Manipulation check. As intended, participants in the human error (vs. algorithm error) condition indicated that the source of the error is more human, $M_{HE} = 5.32$, $SD = 1.41$ vs. $M_{AE} = 4.22$, $SD = 1.73$, $F(1,249) = 30.18$, $p < .001$. Participants in the algorithm error (vs. human error) condition indicated that the source of the error is more algorithm-like, $M_{AE} = 5.07$, $SD = 1.60$ vs. $M_{HE} = 4.03$, $SD = 1.61$, $F(1,249) = 26.33$, $p < .001$.

Brand attitude. A one-way ANOVA on participants' attitude toward the brand, Fiat Chrysler, is significant, $F(1,249) = 4.09$, $p = .044$. Supporting H_1 , participants' responses to the brand following a brand harm crisis are less negative, when the error is an algorithm error (vs. human error), $M_{AE} = 4.59$, $SD = 1.61$ vs. $M_{HE} = 4.17$, $SD = 1.69$ (H_1).

Test of mediation. We next test the mediating role of mind perception of agency of the source of the error in committing the error and the source of the error's responsibility for the harm caused in mediating participants' responses to the brand following a brand harm crisis (H_2). We note that the means of mind perception of agency and responsibility for the harm caused in algorithm error and human error condition are respectively, as follows: $M_{AE} = 3.65$, $SD = 1.65$ vs. $M_{HE} = 4.87$, $SD = 1.37$, $F(1, 249) = 40.66$, $p < .001$; $M_{AE} = 4.53$, $SD = 1.64$ vs. $M_{HE} = 5.11$, $SD = 1.35$, $F(1, 249) = 9.56$, $p = .002$.

We first regressed participants' perceptions of the source of the error's responsibility for the harm caused on algorithm error (vs. human error) condition and found a significant effect, $\beta = .19$, $p = .002$. We then regressed participants' mind perception of source of the error's agency in committing the error on algorithm error (vs. human error) condition and found a significant effect, $\beta = .37$, $p < .001$. We then regressed participants' perception of the source of the error's responsibility for the harm caused on both algorithm error (vs. human error) condition and mind perception of source of the error's agency in committing the error. While there is no effect of

algorithm error (vs. human error) condition, $\beta = .04, p = .54$, there is a significant effect of mind perception of source of the error's agency in committing the error, $\beta = .41, p < .001$.

Next, we formally test the proposed serial mediation model (H_2). We used PROCESS Macro Model 6 (Hayes and Preacher 2014), where algorithm error (vs. human error) are the independent variables, participants' mind perception of source of the error's agency in committing the error and perception of source of the error's responsibility for the harm are the serial mediators, and brand attitude is the dependent variable. The model first tests the effect of the algorithm error (vs. human error) and mind perception of source of the error's agency in committing the error on perception of source of the error's responsibility for the harm caused. The results show no effect of algorithm error (vs. human error) condition, $\beta = .1167, 95\% CI = -.2535$ to $.4869$, but a significant effect of mind perception of source of the error's agency in committing the error on the source of the error's responsibility for the harm caused, $\beta = .3911, 95\% Confidence Interval (CI) = .2762$ to $.5059$.

The model then tests for the effects of algorithm error (vs. human error), mind perceptions of source of the error's agency in committing the error, and perception of source of the error's responsibility for the harm caused on brand attitude. The results show a significant effect of algorithm error (vs. human error) condition ($\beta = -.6306, 95\% CI = -1.0555$ to $-.2058$), participants' mind perception of the source of the error's agency in committing the error ($\beta = .3105, 95\% CI = .1673$ to $.4536$) and perception of the source of the error's responsibility for the harm caused ($\beta = -.2794, 95\% CI = -.4228$ to $-.1360$) on brand attitude. The 95% bias-corrected bootstrap CI for the indirect effect of algorithm error (vs. human error) condition on brand attitude is significant ($\beta = -.1309; 95\% CI = -.2413$ to $-.0498$) indicating serial mediation by

mind perception of source of the error's agency in committing the error and perception of the source of the error's responsibility for the harm caused.

The results of study 2 offer two findings. First, supporting H_1 , consumers' responses to the brand following a brand harm crisis caused by an algorithm (vs. human) error are less negative. Second, in support of H_2 , following a brand harm crisis caused by an algorithm error, participants' mind perception of source of the error's agency in committing the error and perception of source of the error's responsibility for the harm serially mediate consumers' less negative responses to the brand. As the serial mediation is only partial, there may be other theoretical mechanisms that emerge as future research opportunities.

Study 3: Anthropomorphized Algorithm

In study 3, we examine H_3 , that consumers' responses to a brand following a brand harm crisis will be more negative when the error is caused by an anthropomorphized (vs. not) algorithm. We use an incentive compatible experimental design with a consequential outcome, donation to a Feeding America network suggested by the brand, as the dependent variable. We also measured participants' brand attitude. In this study, a fictitious financial investment company, HMS Investments, is facing a crisis because it had made a mistake in the investment decisions of its customers, resulting in financial losses for them. We pre-registered this study on AsPredicted.org (#44090).

Participants and Procedure

Three hundred and seventy-two adults (180 female, $M_{age} = 36.34$, $SD = 14.03$) participated in the experiment on MTurk online platform in exchange for 1 USD. As it is not meaningful to consider anthropomorphized humans, we used a 3-factor experimental design

consisting of algorithm error, human error, and anthropomorphized algorithm error conditions, to which we randomly assign participants. We informed participants that HMS Investments, a leading financial investment company, was facing a crisis because a financial algorithm program (financial manager, or financial algorithm program Charles) had committed an error, resulting in financial losses for its customers.

We measured participants' attitude toward the brand, HMS Investments, using the same five-item scale used in study 2 ($\alpha = .97$). We informed participants that the study's researchers decided to randomly give 20 participants 5\$ bonus, from which participants could donate to Feeding America, the largest domestic hunger-relief organization in the U.S. through HMS Investments. We informed them that each dollar donated provides about 10 meals to families in need through the Feeding America's network of food banks. Participants indicated the amount that they are willing to donate to Feeding America from 0 cents to 500 cents (5\$).

As a manipulation check, we asked participants to indicate the extent to which they thought that the source of the error at HMS Investments was a human and algorithm on a two-item scale (1 = not at all and 7 = very much). Participants also provided perceptions of the extent to which the news was from a credible source and the extent to which the news was believable on a two-item scale (1 = not at all and 7 = very much). Results showed no effect of algorithm error vs. anthropomorphized algorithm error vs. human error conditions on the news' credibility, $F(2,369) = .23, p = .794$, or its believability, $F(2,369) = .653, p = .521$. Finally, participants provided their basic demographic information.

Results and Discussion

Manipulation check. As intended, participants in the human error (vs. algorithm error vs. anthropomorphized algorithm error) condition indicated that the source of the error at HMS

Investments was more human, $M_{HE} = 5.84$, $SD = 1.24$ vs. $M_{AE} = 4.15$, $SD = 1.95$ vs. $M_{AAE} = 4.26$, $SD = 1.81$, $F(2,369) = 38.48$, $p < .001$. As intended, participants in the algorithm error and anthropomorphized algorithm error (vs. human error) conditions indicated that the source of the error at HMS Investment was more algorithm-like, $M_{AE} = 5.40$, $SD = 1.63$ vs. $M_{AAE} = 5.35$, $SD = 1.64$ vs. $M_{HE} = 3.49$, $SD = 1.96$, $F(2,369) = 47.93$, $p < .001$.

Brand attitude. Consistent with H_3 , a one-way ANOVA analysis on participants' brand attitude, HMS Investments is significant, $F(2,369) = 4.19$, $p = .016$. Supporting H_3 , participants' responses to the brand following a brand harm crisis caused by an algorithm error are more negative when the algorithm is anthropomorphized (vs. not), $M_{AAE} = 3.74$, $SD = 1.76$ vs. $M_{AE} = 4.25$, $SD = 1.84$, $p = .027$. Further, in support of H_1 , following a brand harm crisis caused by an algorithm error (vs. human error), participants' brand attitudes are less negative, $M_{AE} = 4.25$, $SD = 1.84$ vs. $M_{HE} = 3.63$, $SD = 1.81$, $p = .008$. Furthermore, participants' brand attitudes are not different for a brand harm crisis caused by an anthropomorphized algorithm error (vs. human error), $M_{AAE} = 3.74$, $SD = 1.76$ vs. $M_{HE} = 3.63$, $SD = 1.81$, $p = .617$.

Donation amount. Consistent with H_3 , a one-way ANOVA analysis on participants' donation is significant, $F(2,369) = 3.18$, $p = .043$. Participants' donations to Feeding America are higher when the brand harm crisis at HMS Investments is caused by an algorithm error (vs. anthropomorphized algorithm error), $M_{AAE} = 160.40$ (cents), $SD = 157.47$ vs. $M_{AE} = 204.98$, $SD = 180.05$, $p = .039$. Following a brand harm crisis, participants' donations to Feeding America are higher than when it is caused by an algorithm error (vs. human error), $M_{AE} = 204.98$, $SD = 180.05$ vs. $M_{HE} = 156.88$, $SD = 165.19$, $p = .029$. Furthermore, participants' donations to Feeding America are not different from when the brand harm crisis is caused by an error caused

by an anthropomorphized algorithm (vs. human), $M_{AAE} = 160.40$, $SD = 157.47$ vs. $M_{HE} = 156.88$, $SD = 165.19$, $p = .864$.

The results of study 3 offer two key findings. First, supporting H_3 , consumers' responses to the brand following a brand harm crisis caused by an algorithm error are more negative when the algorithm is anthropomorphized (vs. not). Second, in support of H_1 , following a brand harm crisis caused by an algorithm error (vs. human error), consumers' responses to the brand are less negative.

Study 4: Machine Learning Algorithm

In study 4, we examine H_4 , that consumers' responses to a brand following a brand harm crisis will be more negative when the error is caused by a machine learning (vs. not) algorithm. In this study, Twitter was facing a crisis because it had made a mistake in the timelines of its users so that some of the displayed tweets had inappropriate and offensive content. We measured participants' attitude toward Twitter following the brand harm crisis.

Participants and Procedure

Three hundred and ten adults participated in the experiment on MTurk online platform in exchange for 1 US dollar (155 male; $M_{age} = 34.95$, $SD = 11.04$). We informed participants that when users log in to Twitter, their home timelines display a stream of tweets from accounts that they have chosen to follow on Twitter.

As in study 3, it is not meaningful to consider machine learning humans. Hence, we again used a 3-factor experimental design consisting of algorithm error, human error, and machine learning algorithm error conditions to which we randomly assigned participants. In the algorithm error (vs. human error) condition, participants read that Twitter uses algorithms (employees) to

evaluate scores and determine which tweets to display. In the machine learning algorithm error condition, participants read that Twitter uses machine learning algorithms to determine which tweets to display. Additionally, participants in the machine learning algorithm error condition read that machine learning algorithms are algorithms that learn from past data and analyses to make their decisions. We then informed participants that there had been some problems in Twitter timelines, which resulted in the incorrect display of tweets for users. Some of these incorrectly displayed tweets had inappropriate content that had offended some Twitter users.

We measured participants' attitude toward Twitter, using the same five item scale used in study 2 ($\alpha = .94$). As a manipulation check, we asked participants their perceptions of whether the source of the error at Twitter was a human (1 = not at all and 7 = very much). Participants also provided perceptions of the extent to which the news was from a credible source and the extent to which the news was believable on a two-item scale (1 = not at all and 7 = very much). Results showed no effect of algorithm error vs. machine learning algorithm error vs. human error conditions on the news' credibility, $F(2,307) = 1.51, p = .22$ or its believability, $F(2,307) = 1.26, p = .28$. Finally, participants indicated whether they had a Twitter account and provided their basic demographic information.

Results and Discussion

Manipulation check. As intended, participants in the human error (vs. algorithm error) condition indicated that the source of the error at Twitter was more human, $M_{HE} = 5.06, SD = 1.75$ vs. $M_{AE} = 4.49, SD = 1.60, p = .014$. There was no significant difference between the participants in the human error (vs. machine learning algorithm error) condition on the extent to which they thought that the source of the error at Twitter was more human, $M_{HE} = 5.06, SD = 1.75$ vs. $M_{MLAE} = 4.79, SD = 1.54, p = .243$. There was also no significant difference between

participants in the algorithm error (vs. machine learning algorithm error) condition on the extent to which they thought that the source of the error at Twitter was more human, $M_{AE} = 4.49$, $SD = 1.60$ vs. $M_{MLAE} = 4.79$, $SD = 1.54$, $p = .176$.

Brand attitude. Consistent with H_4 , a one-way ANOVA analysis on participants' attitude toward Twitter is significant, $F(2,307) = 4.72$, $p = .010$. Supporting H_4 , participants' responses to the brand following a brand harm crisis are more negative when the error was caused by a machine learning (vs. not) algorithm, $M_{MLAE} = 4.21$, $SD = 1.45$ vs. $M_{AE} = 4.76$, $SD = 1.62$, $p = .011$. Further, in support of H_1 , following a brand harm crisis caused by an algorithm error (vs. human error), participants' responses to the brand are less negative, $M_{AE} = 4.76$, $SD = 1.62$ vs. $M_{HE} = 4.20$, $SD = 1.47$, $p = .009$. Following a brand harm crisis caused by a machine learning algorithm (vs. human), participants' responses to the brand are not different, $M_{MLAE} = 4.21$, $SD = 1.45$ vs. $M_{HE} = 4.20$, $SD = 1.47$, $p = .95$. Whether participants have a Twitter account or not did not change the effect of the algorithm error vs. machine learning algorithm error vs. human error) condition on their attitude toward Twitter, $F(2, 306) = 4.58$, $p = .011$.

The results of study 4 offer two key findings. First, supporting H_4 , consumers' responses to the brand following a brand harm crisis are more negative when the error is caused by a machine learning (vs. not) algorithm. Second, in support of H_1 , following a brand harm crisis caused by an algorithm error (vs. human error), consumers' responses to the brand are less negative.

Study 5: Subjective (vs. Objective) Task

In study 5, we test H_5 , that consumers' responses to a brand following a brand harm crisis will be more negative when the algorithm error occurs in a subjective (vs. objective) task. In this study, we informed participants that a leading university in the United States was facing a crisis

because of an error in the subjective (vs. objective) assessment of Asian American students' applications. We measured participants' attitude toward the university. We pre-registered the study at AsPredicted.org (#43396).

Participants and Procedure

Four hundred adults (199 female, $M_{age} = 35.70$, $SD = 11.97$) participated in the study in MTurk online platform in exchange for monetary compensation. We used a 2 (algorithm error, human error) \times 2 (subjective task, objective task) between-subjects design.

We informed participants that a leading university in the U.S. was experiencing a crisis because of a mistake in assessments of the applications of prospective Asian-American students. We randomly assigned participants to either the algorithm error or human error condition. In the algorithm error (human error) condition, we informed participants that the computer algorithm (employees) had made the mistake.

We informed participants that the university used both subjective and objective methods in their admissions process. The subjective methods include analyzing the applicant's personality and social skills including "positive personality," likability, courage, kindness and being "widely respected." The objective methods include reviewing the applicant's test scores and grades. Participants in the subjective (vs. objective) task condition read that the error was in the subjective (vs. objective) assessment of the application. Participants in the subjective task condition read that Asian-American applicants were rated lower than other applicants on traits like "positive personality," likability, courage, kindness and being "widely respected." Participants in the objective task condition read that the error was based on incorrect use of lower test scores for the Asian-American applicants. In both conditions, participants read that the error

1 Author Accepted Manuscript
2
3 resulted in the university incorrectly declining applications of hundreds of Asian-American
4
5 students of otherwise qualified and acceptance-worthy applicants.
6

7
8 We measured participants' attitudes toward the university, using the same five-item scale
9
10 used in study 2 ($\alpha = .95$). As a manipulation check, we asked participants to indicate the extent
11
12 to which they thought that the source of the error was human, the extent to which they thought
13
14 that the source of the error was algorithm, and the extent to which they thought that the error was
15
16 on an objective task (1 = not at all and 7 = very much). Participants also provided perceptions of
17
18 the extent to which the news was from a credible source and the extent to which the news was
19
20 believable (1 = not at all and 7 = very much). Results showed no effect of algorithm error vs.
21
22 believable (1 = not at all and 7 = very much). Results showed no effect of algorithm error vs.
23
24 human error condition and subjective (vs. objective) task conditions on the news' credibility,
25
26 $F(1,396) = .00, p = .994$ or its believability, $F(1,396) = .454, p = .501$. Finally, participants
27
28 provided their basic demographic information.
29

30
31 **Results and Discussion**
32

33
34 *Manipulation check.* As intended, participants in the human error (vs. algorithm error)
35
36 condition indicated that the source of the error was more human, $M_{HE} = 5.33, SD = 1.41$ vs. M_{AE}
37
38 $= 4.45, SD = 1.51, t(398) = 6.003, p < .001$. Participants in the algorithm error (vs. human error)
39
40 condition indicated that the source of the error was more algorithm-like, $M_{AE} = 4.68, SD = 1.64$,
41
42 vs. $M_{HE} = 3.84, SD = 1.68, t(398) = -5.047, p < .001$. Participants in the objective task (vs.
43
44 subjective task) condition indicated that the error occurred is likely to have occurred in an
45
46 objective task, $M_{OBJECTIVE} = 4.60, SD = 1.41$ vs. $M_{SUBJECTIVE} = 3.94, SD = 1.74, t(398) = -4.184$,
47
48 $p < .001$.
49

50
51
52 *Brand attitude.* Consistent with H₅, an ANOVA analysis on participants' brand attitude
53
54 reveals the predicted interaction effect of algorithm error (vs. human error) and subjective (vs.
55
56
57
58
59
60

objective) task conditions, $F(1, 396) = 9.15, p = .003$. Supporting H_5 , participants' responses to a brand following a brand harm crisis caused by an algorithm error are more negative when the error occurs in a subjective (vs. objective) task, $M_{\text{SUBJECTIVE}} = 3.76, SD = 1.64$ vs. $M_{\text{OBJECTIVE}} = 4.46, SD = 1.49, F(1,396) = 9.06, p = .003$. Participants' responses to a brand following a brand harm crisis caused by a human error are not different when the error occurs in a subjective (vs. objective) task, $M_{\text{SUBJECTIVE}} = 4.28, SD = 1.73$ vs. $M_{\text{OBJECTIVE}} = 3.99, SD = 1.64, F(1,396) = 1.58, p = .21$.

In support of H_1 , participants' responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) are less negative when the error occurs in an objective task, $M_{\text{AE}} = 4.46, SD = 1.49$ vs. $M_{\text{HE}} = 3.99, SD = 1.64, F(1,396) = 4.13, p = .043$. Participants' responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) are more negative when the error occurs in a subjective task, $M_{\text{AE}} = 3.76, SD = 1.64$ vs. $M_{\text{HE}} = 4.28, SD = 1.73, F(1,396) = 5.04, p = .025$.

The results of study 5 offer two key findings. First, supporting H_5 , consumers' responses to a brand following a brand harm crisis are more negative when the algorithm error occurs in a subjective (vs. objective) task. Second, supporting H_1 , consumers' responses to the brand following a brand harm crisis caused by an algorithm error (vs. human error) are less negative when the error occurs in an objective task.

Study 6: Interactive Task

In study 6, we test H_6 , that consumers' responses to a brand following a brand harm crisis caused by an algorithm error will be more negative when there is interactivity (vs. not) with the algorithm in the task where the error occurs. We informed participants that a fictitious leading fashion retailer brand, D&J, has been facing growing customer complaints because of their

personal stylists. Participants were randomly assigned to algorithm error (vs. human error) and interactive (vs. non-interactive) task conditions. We measured participants' brand attitude.

Participants and Procedure

Three hundred and twenty-eight students (206 female, $M_{age} = 20.12$, $SD = 1.64$) from a Southern university in the US participated in the laboratory experiment in exchange for course credit. We used a 2 (algorithm error, human error) $\times 2$ (interactive task: yes, no) between-subjects design.

We randomly assigned participants to either the algorithm error or human error conditions. Participants in the algorithm error (vs. human error) condition read that in recent weeks, D&J, a leading fashion retailer brand had been facing growing customer complaints because of some problems caused by its algorithm (human) personal stylists, a recent introduction to personalize products for customers to reflect and accentuate their personalities. Participants were assigned to the interactive (vs. non-interactive) task conditions. To ensure realism, we do not use the word “human” in the human error condition.

Participants in the interactive task condition read that customers who wanted to use the interactive algorithm (personal) stylists, first completed an online form, which collected a personal photograph and details of their height, weight, and personal likes and dislikes of different colors and styles. Then, the D&J algorithm (personal) stylists interact with customers where customers can see how the products will look on them and work with the D&J algorithm (personal) stylists to choose the right products. The customer is thus actively involved in the selection of products by algorithm (personal) stylists. Based on the information provided by the customer, the algorithm (personal) stylists choose and ship products to customers.

Author Accepted Manuscript

Participants in the non-interactive task condition read that customers who use the algorithm (personal) stylists first completed an online form that collected a personal photograph and details of their height, weight, and personal likes and dislikes of different colors and styles. Then, the D&J personal algorithm (personal) stylists choose the right products for the customer. The customer is not involved in the selection of products, done by algorithm (personal) stylists. Based on the information provided by the customer, the algorithm (personal) stylists choose and ship products to customers. The participants read that customers stated that stylists misled them, because of which they had bought very expensive products that did not reflect their personalities and were, in fact, a misfit with their personalities. Customers were now demanding refunds for these products and threatening to sue D&J.

We measured participants' brand attitude, using the same five-item scale used in study 2 ($\alpha = .88$). As a manipulation check, we asked participants to indicate the extent to which they thought that the source of the error at D&J was a human and the extent to which they thought that the error was on a task where there was communication between the personal stylist and the customer, which indicates interactivity on a two-item scale (1 = not at all and 7 = very much). Participants also provided perceptions of the extent to which the news was from a credible source and the extent to which the news was believable on a two-item scale (1 = not at all and 7 = very much). Results showed no effect of algorithm error vs. human error condition and interactive (vs. non-interactive) task conditions on the news' credibility, $F(1,324) = 1.49, p = .22$ or its believability, $F(1,324) = .018, p = .89$. Finally, participants provided their basic demographic information.

Results and Discussion

Manipulation check. As intended, participants in the human error (vs. algorithm error) condition indicated that the source of the error was more human, $M_{HE} = 5.19$, $SD = 1.47$ vs. $M_{AE} = 4.40$, $SD = 1.51$, $t(326) = 4.83$, $p < .001$. Participants in the interactive (vs. non-interactive) task indicated that the error is more likely to have occurred on a task where there was more communication between the personal stylist and the customer, $M_{INTERACTIVE} = 3.51$, $SD = 1.57$ vs. $M_{NON-INTERACTIVE} = 3.04$, $SD = 1.41$, $t(326) = -2.90$, $p = .004$.

Brand attitude. Consistent with H_6 , an ANOVA analysis on brand attitude reveals the predicted interaction effect of algorithm error (vs. human error) and interactive (vs. non-interactive) task conditions, $F(1, 324) = 5.05$, $p = .025$. Supporting H_6 , participants' responses to a brand following a brand harm crisis caused by an algorithm error are more negative when there is interactivity (vs. not) with the algorithm in the task where the error occurs, $M_{INTERACTIVE} = 3.41$, $SD = 1.05$ vs. $M_{NOT} = 3.82$, $SD = 1.22$, $F(1,324) = 6.33$, $p = .012$. Participants' responses to a brand following a brand harm crisis caused by a human error did not differ when there is interactivity (vs. not) with the employee in the task, $M_{INTERACTIVE} = 3.48$, $SD = .98$ vs. $M_{NON-INTERACTIVE} = 3.37$, $SD = .95$, $F(1,324) = .44$, $p = .51$.

In support of H_1 , participants' responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) are less negative when the task where the error occurs is non-interactive, $M_{AE} = 3.82$, $SD = 1.22$ vs. $M_{HE} = 3.37$, $SD = .95$, $F(1,324) = 7.59$, $p = .006$. Participants' responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) are not different when the task is interactive, $M_{AE} = 3.41$, $SD = 1.05$ vs. $M_{HE} = 3.48$, $SD = .98$, $F(1,324) = .179$, $p = .672$.

The results of study 6 offer two findings. First, supporting H_6 , consumers' responses to a brand following a brand harm crisis caused by an algorithm error are more negative when the

task where the error occurs is interactive (vs. non-interactive). Second, supporting H₁, consumers' responses to the brand following a brand harm crisis caused by an algorithm (vs. human) error are less negative when the task where the error occurs is non-interactive.

Managerial Study M1

As algorithm errors are, unfortunately, common in business practice, firms undertake interventions to manage the aftermath of such brand harm crises. The baseline intervention in algorithm errors is technological supervision of the algorithm (e.g., facial recognition algorithm failures at Microsoft) (Roach 2018) to address the algorithm error. Another common intervention following brand harm crises caused by an algorithm error is to increase human supervision of the algorithm (Lee, Resnick, and Barton 2019). As Sheryl Sandberg, Chief Operating Officer, Facebook noted (in 2017) after an algorithm error caused the display of anti-Semitic ads, "we're adding more human review and oversight to our automated processes...From now on we will have more manual review of new ad targeting options to help prevent offensive terms from appearing." To generate managerial guidance, we conducted a study (M1), where we examine consumers' responses to human supervision and technological supervision following brand harm crises caused by an algorithm (vs. human) error.

Participants and Procedure

Three hundred and sixty eight adults (171 female, $M_{\text{age}} = 35.08$, $SD = 11.06$) participated in the study in MTurk online platform in exchange for monetary compensation. We used a 2 (algorithm error, human error) \times 2 (human supervision, technological supervision) between-subjects design. We pre-registered the study at AsPredicted.org (#53178).

Author Accepted Manuscript

All participants saw a tweet on the official Twitter account of the New York Times website announcing the recall of 4.8 million Fiat Chrysler vehicles because of a cruise control problem. We assigned participants to the algorithm (vs. human) error condition. Participants in the algorithm error condition read that the computer algorithm at Fiat Chrysler had made a mistake resulting in a defect in the cruise control system causing a safety hazard. Participants in the human error condition read that the employees of Fiat Chrysler had made a mistake resulting in a defect in the cruise control system causing a safety hazard. We then randomly assigned participants to human supervision (vs. technological supervision) condition. We informed participants in the human supervision condition that Fiat Chrysler would have increased managerial supervision in their manufacturing processes to prevent such errors. We informed participants in the technological supervision condition that Fiat Chrysler would have increased technological supervision in the manufacturing processes to prevent such errors.

We measured participants' attitudes toward the Fiat Chrysler brand using the same five-item scale used in study 2 ($\alpha = .96$). As a manipulation check, we asked participants to indicate the extent to which they thought that the source of the error was human, the extent to which they thought that the source of the error was an algorithm, the extent to which they thought that there will be more human supervision at Fiat Chrysler after defects in the cars, and the extent to which there will be more technological supervision at Fiat Chrysler after defects in the cars on four 7-point scales (1 = not at all and 7 = very much). Participants also provided perceptions of the extent to which the news was believable (1 = not at all and 7 = very much). Results showed no effect of algorithm error vs. human error condition and human supervision (vs. technological) supervision conditions on the news' believability, $F(1,364) = .152, p = .697$. Finally, participants provided their basic demographic information.

Results

Manipulation check. As intended, participants in the human error (vs. algorithm error) condition indicated that the source of the error was more human, $M_{HE} = 5.12$, $SD = 1.54$ vs. $M_{AE} = 4.54$, $SD = 1.75$, $t(366) = -3.41$, $p = .001$. Participants in the algorithm error (vs. human error) condition indicated that the source of the error was more algorithm-like, $M_{HE} = 3.93$, $SD = 1.64$ vs. $M_{AE} = 4.92$, $SD = 1.57$, $t(366) = 5.90$, $p < .001$. Participants in the human supervision (vs. technological supervision) condition indicated, going forward, there will be more human supervision at Fiat Chrysler, $M_{HS} = 5.41$, $SD = 1.44$ vs. $M_{TS} = 5.06$, $SD = 1.47$, $t(366) = 2.28$, $p = .023$. Participants in the technological supervision (vs. human supervision) condition indicated, going forward, there will be more technological supervision at Fiat Chrysler, $M_{HS} = 4.95$, $SD = 1.67$ vs. $M_{TS} = 5.29$, $SD = 1.40$, $t(366) = -2.09$, $p = .037$.

Brand attitude. An ANOVA analysis on brand attitude reveals the predicted interaction effect of algorithm error (vs. human error) and human supervision (vs. technological supervision) conditions, $F(1, 364) = 9.25$, $p = .003$. Participants' responses to a brand following a brand harm crisis caused by an algorithm error are more negative when there is more human supervision (vs. technological supervision), $M_{HS} = 4.13$, $SD = 1.56$ vs. $M_{TS} = 4.71$, $SD = 1.69$, $F(1, 364) = 5.44$, $p = .020$. Participants' responses to a brand following a brand harm crisis caused by a human error are more negative, marginally so, when there is more technological supervision (vs. human supervision), $M_{HS} = 4.63$, $SD = 1.69$ vs. $M_{TS} = 4.15$, $SD = 1.65$, $F(1, 364) = 3.87$, $p = .050$.

Participants' responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) are less negative when there is more technological supervision, $M_{AE} = 4.71$, $SD = 1.69$ vs. $M_{HE} = 4.15$, $SD = 1.65$, $F(1, 364) = 5.27$, $p = .022$. Participants' responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) are more

negative when there is more human supervision, $M_{AE} = 4.13$, $SD = 1.56$ vs. $M_{HE} = 4.63$, $SD = 1.69$, $F(1,364) = 4.02$, $p = .046$.

Study M1’s findings indicate that consumers’ responses to a brand following a brand harm crisis caused by an algorithm error are more (less) negative when there is human (technological) supervision of the algorithm following the harm crisis. The practical implication of these findings is that marketers should not (should) publicize human (technological) supervision of algorithms, when they are used, following brand harm crisis caused by algorithm errors in communications with their customers to ensure superior responses from consumers.

General Discussion

“AI algorithms may be flawed. These deficiencies could undermine the decisions, predictions, or analysis AI applications produce, subjecting us to competitive harm, legal liability, and brand or reputational harm..” Microsoft Annual Report, August 2018.

The use of algorithmic marketing across many applications is growing dramatically across many sectors. Moreover, there is growing evidence of the occurrence of algorithm errors that cause brand harm crises. Yet, there are few insights in the marketing literature on consumers’ responses to brands following a brand harm crisis caused by algorithm errors.

Addressing this research gap, we develop and find support for a theory of consumers’ responses to a brand following a brand harm crisis caused by an algorithm error. The findings from eight experimental studies which support the hypotheses are robust across multiple contexts (e.g., products, financial services, and online services), different samples (e.g., students, adults), and different responses including attitudinal, behavioral, and consequential actions (in two incentive-compatible experimental designs). We conclude with a discussion of the findings’

theoretical contributions, managerial implications, and limitations and opportunities for further research.

Theoretical Contributions

Harm Crises. Distinct from past research on consumers' attributions on product failures caused by managerial (i.e., human) errors, we consider brand harm crises caused by inanimate entities, algorithms which are software programs. Consumers perceive that inanimate algorithms have lower agency over the error and therefore, lower responsibility for the harm caused by the algorithm error.

Applying the theory of mind perception (Gray, Gray, and Wegner 2007) to algorithms that commit errors that cause brand harm crises, we find that consumers have lower mind perception of agency of the algorithm for the error, assign lower responsibility to the algorithm for the harm caused (H_2) resulting in less negative responses to the brand (H_1). Further, consumers' responses to the brand following a brand harm crisis caused by an algorithm error are more negative when 1) the algorithm is anthropomorphized (vs. not) (H_3), 2) it is a machine learning algorithm (vs. not) (H_4), 3) when the algorithm error occurs in a subjective (vs. objective) task (H_5), and 4) when the algorithm error occurs in an interactive (vs. non-interactive) task (H_6).

Taken together, the support for the four moderation effects (i.e., anthropomorphized algorithm, machine learning algorithm, subjective task, and interactive task), each of which humanize the algorithm, indirectly support the serial mediation by lower mind perception of agency of algorithm for the error and in turn, their lower responsibility for the harm caused. Given the growing prevalence of inanimate entities (e.g., algorithms, robots, and drones) in practice, this research's findings make a novel contribution to the literature on harm crisis, which

has not examined consumers’ responses to errors caused by inanimate entities. Further, extant literature has also not examined moderators of the sources of harm crises and characteristics of the task where the error occurs. Finally, the support for partial serial mediation by agency of algorithm for the error and in turn, their lower responsibility for the harm caused by the error suggests that there may be other theoretical processes, which offer future research opportunities.

Algorithm Usage. Extant research on algorithm usage (e.g., Dietvorst et al. 2015, Logg et al. 2019, Prahl and van Swol 2017) has focused on consumers’ decisions to use (or continue to use) an algorithm. However, there may be situations in practice, such as in algorithmic marketing where others, not the algorithm users decide on whether to deploy the algorithm or not. Yet, algorithm errors frequently occur in such contexts, an issue overlooked in extant research. We address this gap and consider consumers’ responses to the brand following a brand harm crisis caused by an algorithm error (vs. human error) where brand managers (not consumers) decide to deploy the algorithm. In what we consider a novel finding, when an algorithm commits an error and causes a brand harm crisis, consumers’ responses to the brand following the crisis are less negative than if the firm’s managers committed the same error. That is, consumers are more forgiving of algorithm errors, suggesting individuals’ receptivity to algorithms when they do not have the decision-making authority on whether to use the algorithm or not.

Further research on individuals’ responses to algorithm errors will be useful, for example, in healthcare, where there is increasing application of algorithms where users do not decide on the usage of the algorithm. For example, in the diagnosis and treatment of health conditions where Big Data are used, there may be the likelihood of different types of errors (e.g., omission or commission, Type I and Type II errors resulting in false positives and false negatives) which may affect consumers’ responses to the brand using the algorithm and to the algorithm itself.

Author Accepted Manuscript

Algorithmic Marketing. To the best of our knowledge, this is the first study to examine consumers' responses to algorithmic errors. We identify consumers' mind perception of agency of algorithms as a building block, relevant in the study of algorithmic marketing. Moreover, the findings of the four moderation effects identify conditions related to error source and task characteristics that modify the main effect of the algorithm error on consumers' responses to the brand. In doing so, we identify building blocks for developing a comprehensive theory of algorithmic marketing. Relevant questions for further research include how consumers may respond to the brand across different algorithm errors in product development, advertising, and targeting settings. A research area with policy implications is the ethicality of algorithmic marketing (e.g., inappropriately targeting/excluding minority identity using facial recognition algorithms) (Spirina 2009).

Managerial Implications

The research's findings from the theory testing offer actionable guidance to managers on the deployment of algorithms in marketing contexts. First, consumers' responses to a brand following a brand harm crisis caused by an algorithm error (vs. human error) are less negative. In addition, consumers' perceptions of the algorithm's lower agency for the error and resultant lower responsibility for the harm caused by the error mediate their responses to a brand following a brand harm crisis caused by an algorithm error. In sum, consumers penalize brands less when an algorithm (vs. human) causes an error that causes a brand harm crisis.

Second, the findings identify conditions where the algorithm appears to be more human consumers' responses to the brand are more negative following a brand harm crisis caused by an algorithm error. Thus, the brand's risk exposure to the harm caused by algorithm error is higher when the algorithm is anthropomorphized (vs. not), it is a machine learning (vs. not) algorithm, it

is used in a subjective (vs. objective) task, or an interactive (vs. non-interactive) task. Marketers must be aware that in contexts where the algorithm appears to be more human, it would be wise to have heightened vigilance in the deployment and monitoring of algorithms and resource allocation for managing the aftermath of brand harm crises caused by algorithm errors.

Third, to manage the aftermath of brand harm crises caused by algorithm errors, managers can highlight the role of the algorithm and the lack of agency of the algorithm for the error, which may attenuate consumers’ negative responses to the brand. However, we caution that highlighting the role of the algorithm will worsen the situation by strengthening consumers’ negative responses for an anthropomorphized algorithm, a machine learning algorithm or if the algorithm error occurs in a subjective or in an interactive task.

Fourth, the insights from the managerial study M1 generate concrete guidance for effectively managing the aftermath of brand harm crises caused by algorithm errors. Marketers should not publicize human supervision of algorithms (which may actually be effective in fixing the algorithm) in communications with their customers following brand harm crisis caused by algorithm errors. However, they should publicize the technological supervision of the algorithm when they use it, to leverage the benefit identified in study M1, i.e., consumers are less negative when there is technological supervision of the algorithm following a brand harm crisis.

Limitations and Further Research

First, in this initial study on brand harm crises caused by algorithm errors, we focus on consumers’ negative responses to brands following one algorithm error. We do not consider the effects of repeated algorithm errors. We also do not consider very serious harm crises with dozens of fatalities (e.g., the Lion Air plane crash in Indonesia in October 2018 and the Ethiopian Airlines plane crash in March 2019 caused by algorithms on Boeing 737 Max 8’s

automated flight system). In such cases, we anticipate extremely negative responses in both algorithm and human error conditions, precluding lab experiments for theory testing. Further, we do not consider marketing mix remedies (e.g., advertising, promotions) that may be effective in handling the aftermath of brand harm crises. Further research on brand harm crises caused by algorithm errors, incorporating marketing mix remedies and their effects on brand performance, using less intrusive, qualitative methods, including observational studies, would be useful. Second, we focus only on errors in algorithmic marketing. Additional research on harm crises caused by algorithmic errors in other contexts (e.g., health care, justice) where algorithm usage is increasing and errors have substantive consequences with policy implications would be useful. Third, with respect to the various parties involved, we consider the brand as the focus of our research without consideration of whether there is a distinction between blaming the algorithm itself, the person who designed it, and the person/company that chose to use it. Future research on whether consumers differentiate between the brand, the designer, the person using the algorithm (e.g., brand manager) emerges as a future research opportunity.

In summary, we view this study as a useful first step in exploring algorithmic marketing, by focusing on brand harm crises caused by algorithm errors that, unfortunately, are now rather common in marketing practice. We hope that this research stimulates further work on algorithmic marketing strategies and related consumer behaviors.

Table 1: Overview of Studies

Study	Participants	Context; Dependent variable	Conditions and Results				Conclusion
Pre-study	N= 403 Online	Financial Investments; Dependent variable (DV): Brand attitude	<u>Error</u>		<u>No Error</u>		When there is no error, consumers' responses to a brand that uses an algorithm (vs. human) are not different. As hypothesized in H ₁ , consumers' responses to a brand following a brand harm crisis caused by an algorithm (vs. human) error are less negative.
			<u>Algorithm</u> 4.55 (1.56) ⁵	<u>Human</u> 3.63 (1.79)	<u>Algorithm</u> 5.55 (1.03)	<u>Human</u> 5.31 (1.07)	
1a	N = 157, Online	Mistake in headline of a fund raising advertisement; DV: Amount of donation	<u>Algorithm Error</u> 20.71 (41.91)		<u>Human Error</u> 7.84 (22.34)		Support for H ₁
1b	N = 233, Online	Online platform had made a mistake and provided wrong financial advice; DV: Advice provided	<u>Algorithm Error</u> 3.19 (2.18)		<u>Human Error</u> 2.49 (1.91)		Support for H ₁
1c	N = 177, U.S., undergrads	Glitch in the online computer system, Qualtrics; DV: % intention to re-engage with the brand	<u>Algorithm Error</u> 64.3%		<u>Human Error</u> 42.1%		Support for H ₁
2	N = 251, Online	Recall of 4.8 million vehicles; DV: Brand Attitude Mediators: mind perception of source of the error's agency in committing the error; perceptions of source of the error's responsibility for the harm caused to the brand.	<u>Algorithm Error</u> 4.59 (1.61)		<u>Human Error</u> 4.17 (1.69)		Support for H ₁ and H ₂ , of serial mediation by lower agency of the algorithm and responsibility for the harm caused by the algorithm error

⁵ Figures in parentheses are standard deviations.

3	N = 372; Online	Mistake in investment decisions of customers of a financial investment company; DV: Brand attitude; Donation amount	<u>Human Error</u> 3.63 (1.81) 156.88 (165.19)		<u>Algorithm Error</u> 4.25 (1.84) 204.98 (180.05)	<u>Anthropomorphized Algorithm Error</u> 3.74 (1.76) 160.40 (157.47)	As hypothesized in H ₃ , consumers' responses to a brand following a brand harm crisis are more negative when the error is caused by an anthropomorphized (vs. not) algorithm.
4	N = 310; Online	Mistake in Twitter timelines of users ; DV: Brand attitude	<u>Human Error</u> 4.20(1.47)		<u>Algorithm Error</u> 4.76 (1.62)	<u>Machine Learning Algorithm Error</u> 4.21 (1.45)	As hypothesized in H ₄ , consumers' responses to a brand following a brand harm crisis are more negative when the error is caused by a machine learning (vs. not) algorithm.
5	N =400, Online	A leading university in the United States was facing a crisis because of an error in the subjective (vs. objective) assessment of Asian American students' applications DV: Brand attitude	<u>Algorithm Error</u> <u>Subjective Task</u> 3.76 (1.64)		<u>Human Error</u> <u>Subjective Task</u> 4.28 (1.73)		As hypothesized in H ₅ , consumers' responses to a brand following a brand harm crisis are more negative when the algorithm error occurs in a subjective (vs. objective) task.
6	N =328, US undergraduate students	Mistake in product selection by a personal stylist of a fashion retailer company DV: Brand attitude	<u>Algorithm Error</u> <u>Interactive Task</u> 3.41 (1.05)		<u>Human Error</u> <u>Non-interactive Task</u> 3.48 (0.98)		Support for H ₆ , consumers' responses to a brand following a brand harm crisis caused by an algorithm error are more negative when the error occurs in an interactive (vs. non-interactive) task.
M1	N = 368, Online	Recall of 4.8 million vehicles DV: Brand attitude	<u>Algorithm Error</u> <u>Technological supervision</u> 4.71 (1.69)		<u>Human Error</u> <u>Human supervision</u> 4.15 (1.65)		Responses to a brand following a brand harm crisis caused by an algorithm error are more negative when there is more human supervision (vs. technological supervision)

References

- Aggarwal, Pankaj and Ann L. McGill (2007), "Is That Car Smiling at Me? Schema Congruity as a Basis for Evaluating Anthropomorphized Products," *Journal of Consumer Research*, 34, 468-479.
- , and ---- (2012), "When Brands Seem Human, Do Humans Act Like Brands? Automatic Behavioral Priming Effects of Brand Anthropomorphism," *Journal of Consumer Research*, 39(2), 307-323.
- Ahluwalia, Rohini, Robert E. Burnkrant, and H. Rao Unnava (2000), "Consumer Response to Negative Publicity: The Moderating Role of Commitment," *Journal of Marketing Research*, 37 (2), 203-214.
- Awad, Edmond, et al (2020), "Drivers are Blamed More than their Automated Cars when Both Make Mistakes," *Nature Human Behaviour*, 4(2), 34-143.
- Badger, Emily, (2019), "Who's to Blame when Algorithms Discriminate at <https://www.nytimes.com/2019/08/20/upshot/housing-discrimination-algorithms-hud.html> accessed on September 20, 2020.
- Castelo, Noah, Maarten W. Bos, and Donald R. Lehmann (2019), "Task-Dependent Algorithm Aversion," *Journal of Marketing Research*, 56(5), 809-825.
- Choi, Sungwoo, Anna S. Mattila, and Lisa E. Bolton (2021), "To Err is Human(-oid): How Do Consumers React to Robot Service Failure and Recovery?", *Journal of Service Research* (forthcoming).
- Cleeren, Kathleen, Marnik G. Dekimpe, and Harald van Heerde (2017), "Marketing Research on Product-Harm Crises: A Review, Managerial Implications, and an Agenda for Future Research," *Journal of the Academy of Marketing Science*, 45(5), 593-615.
- , ----, and Kristiaan Helsen (2008), "Weathering Product-Harm Crises," *Journal of the Academy of Marketing Science*, 36 (2), 262-270.
- Diakopoulos, Nicholas (2013), "Race Against the Algorithms," *The Atlantic*, <https://www.theatlantic.com/technology/archive/2013/10/rage-against-the-algorithms/280255/>
- Dietvorst, Berkeley, Joseph P. Simmons, and Cade Massey (2015), "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err," *Journal of Experimental Psychology: General*, 144 (1), 114-126.
- Dutta, Sujay and Chris Pullig (2011), "Effectiveness of Corporate Responses to Brand Crises: The Role of Crisis Type and Response Strategies," *Journal of Business Research*, 64 (12), 1281-1287.
- Epley, Nicholas and Adam Waytz (2009), Mind Perception. In S.T. Fiske, D.T. Gilbert, and G. Lindzey (Eds.), *The Handbook of Social Psychology*, 5th edition. New York, New York: Wiley.
- Epley, Nicholas, Eugene Caruso, and Max H. Bazerman (2006), "When Perspective Taking Increases Taking: Reactive Egoism in Social Interaction," *Journal of Personality and Social Psychology*, 91, 872.
- Fletcher, Joseph F. (1979), *Humanhood: Essays in Biomedical Ethics*. Buffalo, New York: Prometheus Books

Author Accepted Manuscript

- Folkes, Valerie S. (1984), "Consumer Reactions to Product Failure: An Attributional Approach," *Journal of Consumer Research*, 10 (4), 398-409.
- (1990). *Conflict in the Marketplace: Explaining Why Products Fail*. In S. Graham and Valerie S. Folkes eds., *Attribution Theory: Applications to Achievement, Mental Health, and Interpersonal Conflict*. Hillsdale, New Jersey: Lawrence Erlbaum.
- , Susan Koletsky, and John L. Graham (1987), "A Field Study of Causal Inferences and Consumer Reaction: The View from the Airport," *Journal of Consumer Research*, 15, 534-539.
- Gal, Michal S. and Niva Elkin-Koren (2017), "Algorithmic Consumers," *Harvard Journal of Law & Technology*, 30 (2), 309-353.
- Gill, Tripat (2020), "Blame It on the Self-Driving Car: How Autonomous Vehicles Can Alter Consumer Morality," *Journal of Consumer Research*, 47(2), 272-291.
- Gray, Heather M., Kurt Gray, and Daniel M. Wegner (2007), "Dimensions of Mind Perception," *Science*, 315, 619.
- Gray, Kurt and Daniel M. Wegner (2009), "Moral Typecasting: Divergent Perceptions of Moral Agents and Moral Patients," *Journal of Personality and Social Psychology*, 96(3), 505-520.
- and ---- (2012), "Feeling Robots and Human Zombies: Mind Perception and the Uncanny Valley," *Cognition*, 125(1), 125-130.
- , Liane Young, and Adam Waytz (2012), "Mind Perception Is the Essence of Morality," *Psychological Inquiry*, 23(2), 101-124.
- Griffith, Eric (2017), "10 Embarrassing Algorithm Fails," *PCMag*, <https://www.pcmag.com/feature/356387/10-embarrassing-algorithm-fails> accessed on March 10, 2019.
- Hayes, Andrew F. and Kristopher J. Preacher (2014), "Statistical Mediation Analysis with a Multicategorical Independent Variable," *British Journal of Mathematical and Statistical Psychology*, 67, 451-470.
- Heller, Martin (2019), "Machine learning algorithms explained," *InfoWorld*, <https://www.infoworld.com/article/3394399/machine-learning-algorithms-explained.html>.
- Inbar, Yoel, Jeremy Cone, and Thomas Gilovich (2010), "People's Intuitions about Intuitive Insight and Intuitive Choice," *Journal of Personality and Social Psychology*, 99(2), 232-247.
- Kim, Sara and Ann McGill (2011), "Gaming with Mr. Slot or Gaming the Slot Machine? Power, Anthropomorphism, and Risk Perception," *Journal of Consumer Research*, 38(1), 94-107.
- Kim, Hyeongmin Christian and Thomas Kramer (2015), "Do Materialists Prefer the "Brand-as-Servant"? The Interactive Effect of Anthropomorphized Brand Roles and Materialism on Consumer Responses," *Journal of Consumer Research*, 42(2), 284-299.
- Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018), "Human Decisions and Machine Predictions," *The Quarterly Journal of Economics*, 133 (1), 237-293.

- Lafrance, Adrienne (2014), "Why People Name Their Machines," *The Atlantic*, <https://www.theatlantic.com/technology/archive/2014/06/why-people-give-human-names-to-machines/373219/>.
- Landwehr, Jan R., Ann McGill, and Andreas Herrmann (2011), "It's Got the Look: The Effect of Friendly and Aggressive "Facial" Expressions on Product Liking and Sales," *Journal of Marketing*, 75(3), 132-146.
- Lee, Nicol Turner, Paul Resnick, and Genie Barton (2019), "Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms," *Brookings*, <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>
- Lei, Jing, Niraj Dawar, and Zeynep Gürhan-Canli (2012), "Base-Rate Information in Consumer Attributions of Product-Harm Crises," *Journal of Marketing Research*, 49 (3), 336-348.
- Logg, Jennifer, Julia Minson, and Don A. Moore, (2019), "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment," *Organizational Behavior and Human Decision Processes*, 151, 90-103.
- McCullom, Rod (2017), "Facial Recognition Technology Is Both Biased and Understudied" <https://undark.org/article/facial-recognition-technology-biased-understudied/>
- Moon, Youngme (2000), "Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers," *Journal of Consumer Research*, 26(4), 323-339.
- (2003), "Don't Blame the Computer: When Self-Disclosure Moderates the Self-Serving Bias," *Journal of Consumer Psychology*, 13(1-2), 125-137.
- Nass, Clifford, and Youngme Moon (2000), "Machines and Mindlessness: Social Responses to Computers," *Journal of Social Issues*, 56(1), 81-103.
- Pullig, Chris, Richard G. Netemeyer, and Abhijit Biswas (2006), "Attitude Basis, Certainty, and Challenge Alignment: A Case of Negative Brand Publicity," *Journal of the Academy of Marketing Science*, 34 (4), 528-542.
- Prahl, Andrew, and Lyn Van Swol (2017), "Understanding Algorithm Aversion: When is Advice from Automation Discounted?" *Journal of Forecasting*, 36(6), 691-702.
- Puzakova, M., Kwak, H. and Rocereto, J.F. (2013), "When Humanizing Brands Goes Wrong: The Detrimental Effect of Brand Anthropomorphization Amid Product Wrongdoings". *Journal of Marketing*, 77, 81-100.
- Rafaeli, Sheizaf (1988), "Interactivity: From New Media to Communication," In *Sage Annual Review of Communication Research: Advancing Communication Science*, Vol. 16, eds R. P. Hawkins, J. M. Wiemann and S. Pingree, 110– 134. Beverly Hills , CA : Sage.
- Roach, John (2018), "Microsoft Improves Facial Recognition Technology to Perform Well Across All Skin Tones , Genders," *Microsoft The AI Blog*, <https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/>
- Sandberg, Sheryl (2017), <https://www.facebook.com/sheryl/posts/10159255449515177>
- Spirina, Katrine (2009), "Ethics of Facial Recognition: How to Make Business Uses Fair and Transparent," *Towards Data Science*, <https://towardsdatascience.com/ethics-of-facial-recognition-how-to-make-business-uses-fair-and-transparent-98e3878db08d> accessed on January 4, 2020.

Author Accepted Manuscript

- Sundar, S. S. (2009), "Media Effects 2.0: Social and Psychological Effects of Communication Technologies," In R. L. Nabi, & M. B. Oliver (Eds.). *The SAGE Handbook of Media Processes and Effects*. Thousand Oaks, CA: Sage Publications.
- , Saraswathi Bellur, Jeeyun Oh, Haiyan Jia, and Hyang Sook Kim (2016), "Theoretical Importance of Contingency in Human-Computer Interaction: Effects of Message Interactivity on User Engagement," *Communication Research*, 43(5), 595-625.
- , Eun Go, Hyang-Sook Kim, and Bo Zhang (2015), "Communicating Art, Virtually! Psychological Effects of Technological Affordances in a Virtual Museum," *International Journal of Human-Computer Interaction*, 31(6), 385-401.
- Swaminathan, Vanitha, Karen L. Page, and Zeynep Gürhan-Canli (2007), "'My' Brand or 'Our' Brand: The Effects of Brand Relationship Dimensions and Self-Construal on Brand Evaluations," *Journal of Consumer Research*, 34, 248-259.
- Sweeney, Latanya (2013), "Discrimination in Online Ad Delivery," *Queue*, 56(5), 10.
- Vigdor, Neil (2019), "Apple Card Investigated after Gender Discrimination Complaints," *The New York Times*, <https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html>
- Vincent, James (2019), "Google and Microsoft Warn Investors That Bad AI Could Harm Their Brand," *The Verge*, <https://www.theverge.com/2019/2/11/18220050/google-microsoft-ai-brand-damage-investors-10-k-filing>
- Waytz, Adam, John Cacioppo, and Nicholas Epley (2010), "Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism," *Perspectives on Psychological Science*, 5(3), 219-232.
- , Joy Heafner, and Nicholas Epley (2014), "The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle," *Journal of Experimental Social Psychology*, 52, 113-117.
- and Liane Young (2012), "The Group-Member Mind Tradeoff: Attributing Mind to Groups versus Group Members," *Psychological Science*, 23, 77-85.

Web Appendix: Additional Study and Stimuli for All Studies

Study 1c

In study 1c, we examine consumers’ responses to a brand harm crisis caused by a failure in the online computer system, Qualtrics, that participants used in the laboratory setting, that was caused either by an algorithm error (vs. human error). We randomly assigned participants either to the algorithm error or human error condition and noted their decision to repeat or not repeat the online task (i.e. re-engage with Qualtrics), a behavioral measure of participants’ responses to the brand following a brand harm crisis. Participants’ willingness to repeat the task would indicate a less negative response to the Qualtrics brand.

Participants and Procedure

The experiment used the algorithm error (vs. human error) condition as a between-subjects design. One hundred and eighty-four students from a Southern university in the US participated in the experiment in exchange for course credit. We excluded from the analysis, seven participants who copied and pasted the study’s instructions as their only responses. We conducted the analyses with data from 177 participants (93 male; $M_{age} = 20.72$, $SD = 2.42$).

We instructed participants to transcribe a scientific article from the New York Times newspaper on their computers. On completing the writing task, the lab administrator informed participants that the computer had not recorded their responses because of a glitch caused by an algorithm error (vs. human error). The lab administrator then asked the participants whether they were willing to repeat the task. Participants then provided their basic demographic information. We debriefed the participants, all of whom received full course credit for participation.

Results and Discussion

Author Accepted Manuscript

Re-engage with the brand. There is a significant effect of algorithm error (vs. human error) condition on participants' decision to repeat the task ($\chi^2 = 8.367, p = .004$). 64.3% of the participants in the algorithm error condition chose to repeat the task, while only 42.1% of the participants in the human error condition did so. The results support our prediction (H_1) that consumers' re-engagement behaviors with the brand, following a brand harm crisis caused by an algorithm error (vs. human error) are less negative.

Peer Review Version

Pre-study – Stimuli

Algorithm and no error condition



ENGLISH ESPAÑOL 中文

SUBSCRIBE NOW

LOG IN

Tuesday, June 23, 2020

The New York Times

Today's Paper

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

HMS Investments Reduces Risks for Its Clients With Its Strong Computer Algorithms



Audrey Sanders

HMS Investments is a leading financial investment company. HMS Investments, which has been active since 1955, manages stocks and investments. Operating in multiple cities around the world, HMS Investments invests money on behalf of its clients and has a strong reputation for having strong computer algorithms that reduces risks for its clients seeking to buy stocks in different countries for their investment portfolios.

Algorithm and error condition



ENGLISH ESPAÑOL 中文

SUBSCRIBE NOW

LOG IN

Tuesday, June 23, 2020

The New York Times

Today's Paper

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

HMS Investments Faces a Crisis



Audrey Sanders

HMS Investments, a leading financial investment company is facing a crisis. HMS Investments, which has been active since 1955, manages stocks and investments. Operating in multiple cities around the world, HMS Investments invests money on behalf of its clients and has a strong reputation for reducing risks for its clients seeking to buy stocks in different countries for their investment portfolios.

On June 22, 2020, HMS Investments informed its customers that a financial algorithm program used by HMS Investments had made a mistake, resulting in losses, for its customers. For some of HMS Investment’s customers, the losses were to the extent of hundreds of thousands of dollars.

Human and no error condition**HMS Investments Reduces Risks for Its Clients With Its Strong Employees**

Audrey Sanders

HMS Investments is a leading financial investment company. HMS Investments, which has been active since 1955, manages stocks and investments. Operating in multiple cities around the world, HMS Investments invests money on behalf of its clients and has a strong reputation for having a strong employee workforce that reduces risks for its clients seeking to buy stocks in different countries for their investment portfolios.

Human and error condition**HMS Investments Faces a Crisis**

Audrey Sanders

HMS Investments, a leading financial investment company is facing a crisis. HMS Investments, which has been active since 1955, manages stocks and investments. Operating in multiple cities around the world, HMS Investments invests money on behalf of its clients and has a strong reputation for reducing risks for its clients seeking to buy stocks in different countries for their investment portfolios.

On June 22, 2020, HMS Investments informed its customers that a financial manager at HMS Investments had made a mistake, resulting in losses, for its customers. For some of HMS Investment's customers, the losses were to the extent of hundreds of thousands of dollars.

Study 1a – Stimuli
Human error condition:

Forbes

BillionsairesInnovationLeadershipMoneyBusinessSmall BusinessLifestyleListsAdvisorFeaturedBreakingMore



Audrey Sanders

AI & Big Data

I write about the broad intersection of data and society.

Qualtronics, faces a crisis because of a fund raising campaign, BanishCovid19, aimed at combating Covid 19, which they implied was caused by a Chinese Virus

f

Qualtronics, a leading consumer electronics retailer, is facing a crisis. On March 25, 2020, Qualtronics started a fund raising campaign to help combat COVID 19 caused by the Coronavirus.

tw

The funds from this campaign would go to the World Health Organization. Because the disease was first detected in Wuhan Province in China, Qualtronics' managers designed the advertisement and released a campaign with the headline

in

"Contribute to BanishCovid19 and Destroy the Chinese Virus."

Following negative feedback from their customers on this advertisement, Qualtronics apologized to its customers and changed the advertisement headline to "Contribute to BanishCovid19 and Destroy the Corona Virus."

Algorithm error condition:

Forbes

BillionsairesInnovationLeadershipMoneyBusinessSmall BusinessLifestyleListsAdvisorFeaturedBreakingMore



Audrey Sanders

AI & Big Data

I write about the broad intersection of data and society.

Qualtronics, faces a crisis because of a fund raising campaign, BanishCovid19, aimed at combating Covid 19, which they implied was caused by a Chinese Virus

f

Qualtronics, a leading consumer electronics retailer, is facing a crisis. On March 25, 2020, Qualtronics started a fund raising campaign to help combat COVID 19 caused by the Coronavirus.

tw

The funds from this campaign would go to the World Health Organization. Because the disease was first detected in Wuhan Province in China, Qualtronics which uses computer algorithms to develop its advertisements released a campaign with the headline

in

"Contribute to BanishCovid19 and Destroy the Chinese Virus."

Following negative feedback from their customers on this advertisement, Qualtronics apologized to its customers and changed the advertisement headline to "Contribute to BanishCovid19 and Destroy the Corona Virus."

Study 1b – Stimuli**Human error condition:**

Life Skills without Borders is a global platform company operating all around the world. It commits to provide basic life skills for all children and youth. They gather information from people with life experiences and share their experiences with under-privileged children and youth to develop the life skills that they may need in their lives.

Life Skills without Borders is experiencing a crisis because an employee made a mistake such that incorrect financial advice was provided to young poor couples. Specifically, these young couples with limited financial resources were advised to invest in risky company stocks, which may have resulted in them losing their money. The extent of their financial losses is still under investigation.

Life Skills without Borders is now crowdsourcing ideas for providing life skills advice to first generation to college, young adults looking for their first job after graduation.

Based on your life experiences, please provide as much advice as you would like to share with these young adults looking for a job. Please provide vivid detailed information in the form of bullet points.

Algorithm error condition:

Life Skills without Borders is a global platform company operating all around the world. It commits to provide basic life skills for all children and youth. They gather information from people with life experiences and share their experiences with under-privileged children and youth to develop the life skills that they may need in their lives.

Life Skills without Borders is experiencing a crisis because a computer algorithm made a mistake such that incorrect financial advice was provided to young poor couples. Specifically, these young couples with limited financial resources were advised to invest in risky company stocks, which may have resulted in them losing their money. The extent of their financial losses is still under investigation.

Life Skills without Borders is now crowdsourcing ideas for providing life skills advice to first generation to college, young adults looking for their first job after graduation.

Based on your life experiences, please provide as much advice as you would like to share with these young adults looking for a job. Please provide vivid detailed information in the form of bullet points.

Study 1c – Stimuli

All participants wrote the following article:

Even a single workout could be good for the heart. That’s the conclusion of a fascinating new study in mice that found that 30 minutes on a treadmill affects gene activity within cardiac cells in ways that, over the long haul, could slow the aging of the animals’ hearts.

Although the study involved mice, the results may help to explain just how, at a cellular level, exercise improves heart health in people as well.

There’s no question that, in general, physical activity is good for hearts. Many studies have found that people who regularly exercise are much less likely to develop or die from cardiac disease than people who are sedentary.

Still, researchers have remained puzzled about just how exercise alters hearts for the better. Exercise is known to improve our blood pressure, pulse rate and cholesterol profiles, all of which are associated with better cardiac health.

But many scientists who study the links between exercise and heart health have pointed out that these changes, considered together, explain only about half of the reported statistical reductions in cardiac disease and death.

Other, more complex physiological modifications must simultaneously be taking place within the heart itself during and after exercise, these researchers have speculated. And recently, researchers at the University of Maryland in College Park and other institutions have begun to wonder whether some of these changes might involve telomeres.

Telomeres are tiny caps on the ends of chromosomes, often compared to the tips of shoelaces, which help to prevent fraying and damage to our DNA. Young cells have relatively long telomeres. As a cell ages or undergoes significant stress, its telomeres shorten. If they become too abbreviated, the cell stops working well or dies.

But while shorter telomeres indicate biologically older cells, the process is not strictly chronological, scientists have found. Cells can age at different rates, depending on the lifestyle of the body that contains them.

Aerobic exercise, in particular, affects telomeres. In past studies, masters athletes have been shown to have longer telomeres in their white blood cells than sedentary people of the same chronological years, suggesting that at a cellular level, the athletes are more youthful.

But while it is easy to obtain and look inside white blood cells, far less has been known about telomeres within cardiac cells.

So for the new study, which was published this month in Experimental Physiology, the Maryland researchers and their colleagues turned to young, healthy female mice. (They chose females because they tend to run more readily than males.)

Algorithm error condition:

Author Accepted Manuscript

We're sorry. There has been a glitch in the system of Qualtrics, which is caused by an algorithmic error so that your responses could not be recorded. You will have to do the task again.

Human error condition:

We're sorry. There has been a glitch in the system of Qualtrics, which is caused by the managers working at Qualtrics so that your responses could not be recorded. You will have to do the task again.

Peer Review Version

Study 2 Stimuli



Human error condition:

Fiat Chrysler recalls 4.8 million US cars

Fiat Chrysler is recalling 4.8 million US vehicles over a defect that could prevent drivers from turning off cruise control. It warned owners not to use the function until they get software upgrades.

Most of the vehicles being recalled cover models built between 2014-2018. The spokesman of Fiat Chrysler noted that no injuries or crashes were related because of this product recall.

Cruise control system was developed by employees of Fiat Chrysler. Fiat Chrysler employees had made a mistake resulting in a defect in the cruise control system causing a safety hazard. Fiat Chrysler issued a statement that "The company will inform buyers of these defective cars through a registered letter".

Algorithm error condition:

Fiat Chrysler recalls 4.8 million US cars

Fiat Chrysler is recalling 4.8 million US vehicles over a defect that could prevent drivers from turning off cruise control. It warned owners not to use the function until they get software upgrades.

Most of the vehicles being recalled cover models built between 2014-2018. The spokesman of Fiat Chrysler noted that no injuries or crashes were related because of this product recall.

Cruise control system was developed by a computer algorithm at Fiat Chrysler. The computer algorithm at Fiat Chrysler had made a mistake resulting in a defect in the cruise control system causing a safety hazard. Fiat Chrysler issued a statement that "The company will inform buyers of these defective cars through a registered letter".

Study 3 – Stimuli

Author Accepted Manuscript

Human error condition:



HMS Investments, a leading financial investment company is facing a crisis. HMS Investments, which has been active since 1955, manages stocks and investments. Operating in multiple cities around the world, HMS Investments invests money on behalf of its clients and has a strong reputation for reducing risks for its clients seeking to buy stocks in different countries for their investment portfolios.

On January 13, 2020, HMS Investments informed its customers that a financial manager at HMS Investments had made a mistake, resulting in losses, for its customers. For some of HMS Investment's customers, the losses were to the extent of hundreds of thousands of dollars.

Algorithm error condition:



HMS Investments, a leading financial investment company is facing a crisis. HMS Investments, which has been active since 1955, manages stocks and investments. Operating in multiple cities around the world, HMS Investments invests money on behalf of its clients and has a strong reputation for reducing risks for its clients seeking to buy stocks in different countries for their investment portfolios.

On January 13, 2020, HMS Investments informed its customers that a financial algorithm program used by HMS Investments had made a mistake, resulting in losses, for its customers. For some of HMS Investment's customers, the losses were to the extent of hundreds of thousands of dollars.

Anthropomorphized algorithm error condition:



HMS Investments Faces a Crisis



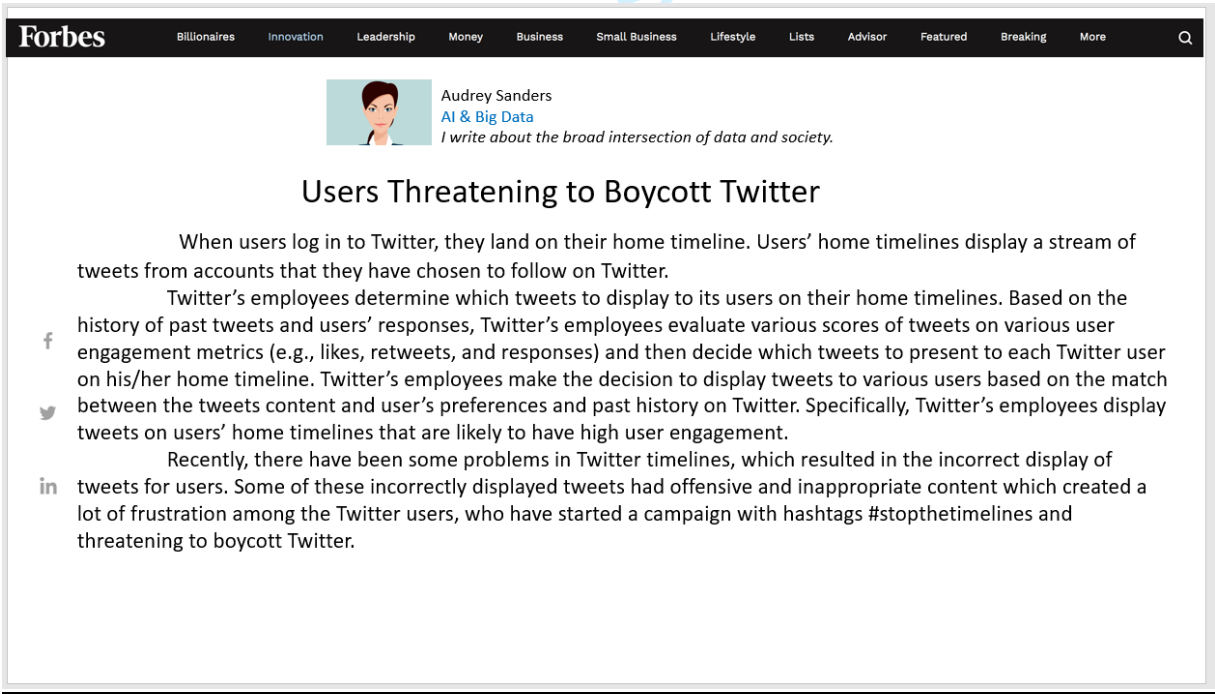
By Adam Liptak

HMS Investments, a leading financial investment company is facing a crisis. HMS Investments, which has been active since 1955, manages stocks and investments. Operating in multiple cities around the world, HMS Investments invests money on behalf of its clients and has a strong reputation for reducing risks for its clients seeking to buy stocks in different countries for their investment portfolios.

On January 13, 2020, HMS Investments informed its customers that a financial algorithm program, Charles, at HMS Investments had made a mistake, resulting in losses, for its customers. For some of HMS Investment’s customers, the losses were to the extent of hundreds of thousands of dollars.

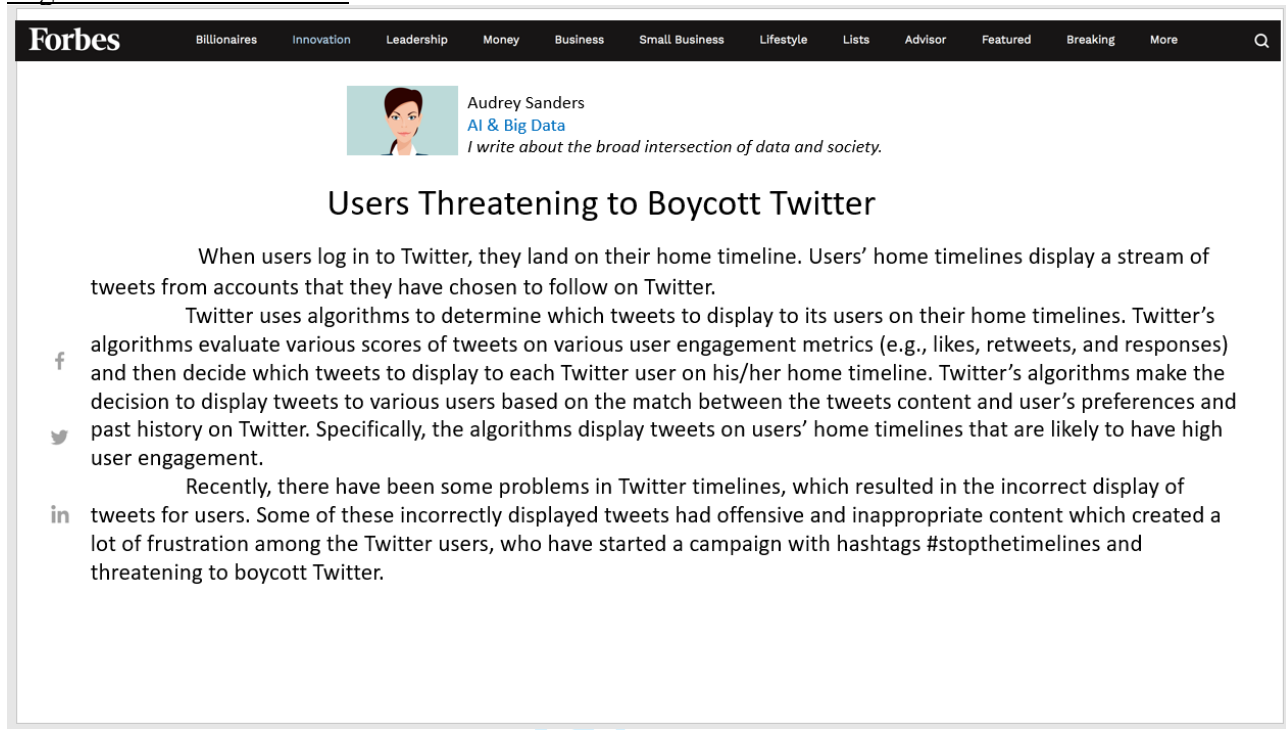
Study 4 Stimuli

Human error condition:

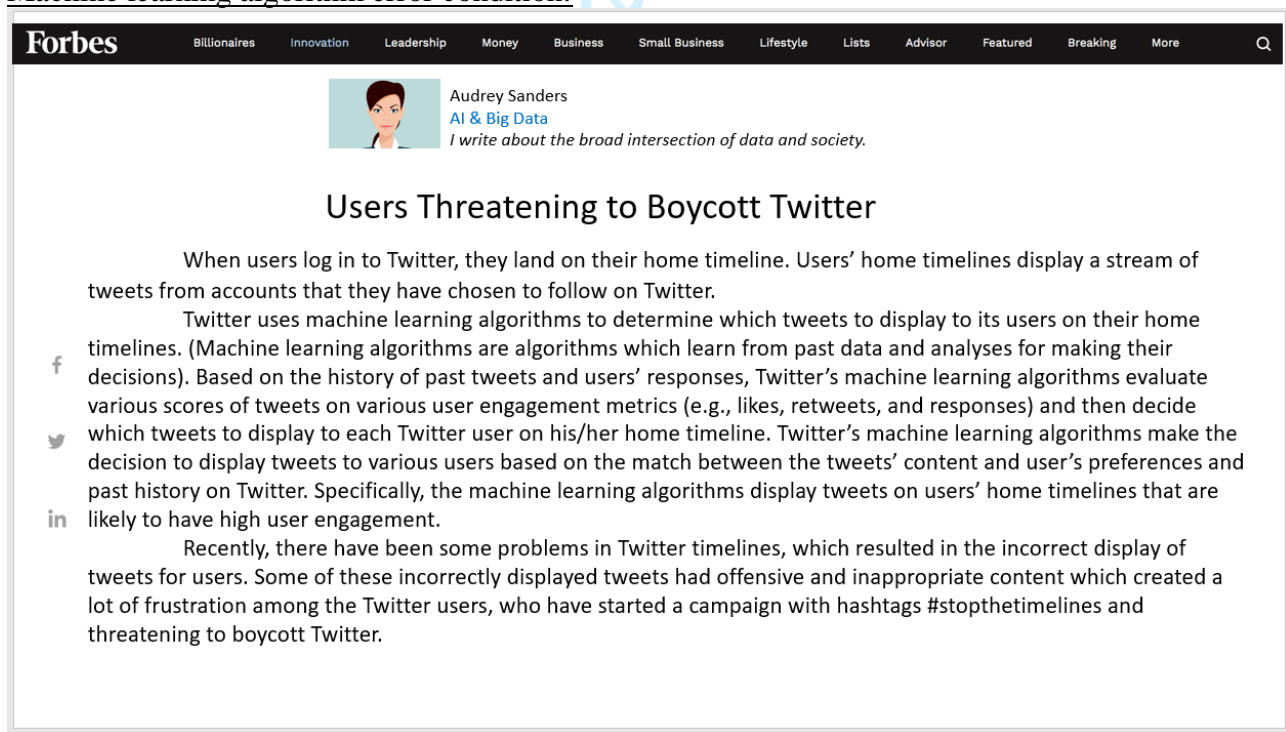


Author Accepted Manuscript

Algorithm error condition:



Machine learning algorithm error condition:



Study 5 - Stimuli

Human error in objective task condition:

SHARE



ENGLISH ESPAÑOL 中文

The New York Times

Monday, June 22, 2020

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

SUBSCRIBE NOW

LOG IN

Today's Paper

Leading Private University Rated Asian-American Applicants Lower on Test Scores, Suit Says

By [Audrey Sanders](#)
Updated June 22, 2020 7:01 pm ET

A leading private university in the United States, admitted only 4.6 percent of its applicants this year. That has led to intense interest in the university’s closely guarded admissions process.

The university used an in-house approach for the admissions process, which include both subjective and objective methods. The objective methods include reviewing the applicant’s test scores and grades. The subjective methods include analyzing the applicant’s traits like “positive personality,” likability, courage, kindness and being “widely respected.”

The university is now is facing a crisis because their employees had made a mistake in their objective assessment of student applications. The employees had incorrectly used lower test scores for the Asian-American applicants. This resulted in declining the applications of hundreds of Asian-American students of otherwise acceptance-worthy applicants. Now, more than 160,000 student records filed by a group representing Asian-American students in a lawsuit against the university.

Human error in subjective task condition:

SHARE



ENGLISH ESPAÑOL 中文

The New York Times

Monday, June 22, 2020

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

SUBSCRIBE NOW

LOG IN

Today's Paper

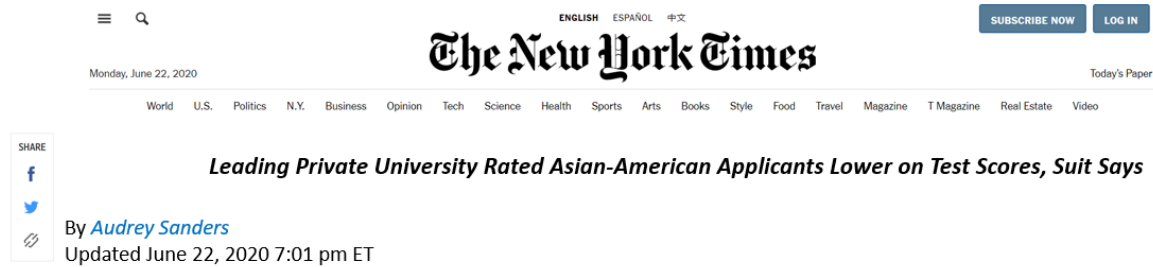
Leading Private University Rated Asian-American Applicants Lower on Personality Traits, Suit Says

By [Audrey Sanders](#)
Updated June 22, 2020 7:01 pm ET

A leading private university in the United States, admitted only 4.6 percent of its applicants this year. That has led to intense interest in the university’s closely guarded admissions process.

The university used an in-house approach for the admissions process, which include both subjective and objective methods. The objective methods include reviewing the applicant’s test scores and grades. The subjective methods include analyzing the applicant’s traits like “positive personality,” likability, courage, kindness and being “widely respected.”

The university is now is facing a crisis because their employees had made a mistake in their subjective assessment of student applications. The employees had incorrectly rated Asian-American applicants lower than others on traits like “positive personality,” likability, courage, kindness and being “widely respected.” This resulted in declining the applications of hundreds of Asian-American students of otherwise acceptance-worthy applicants. Now, more than 160,000 student records filed by a group representing Asian-American students in a lawsuit against the university.

Algorithm error in objective task condition:


Monday, June 22, 2020

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

SHARE

Leading Private University Rated Asian-American Applicants Lower on Test Scores, Suit Says

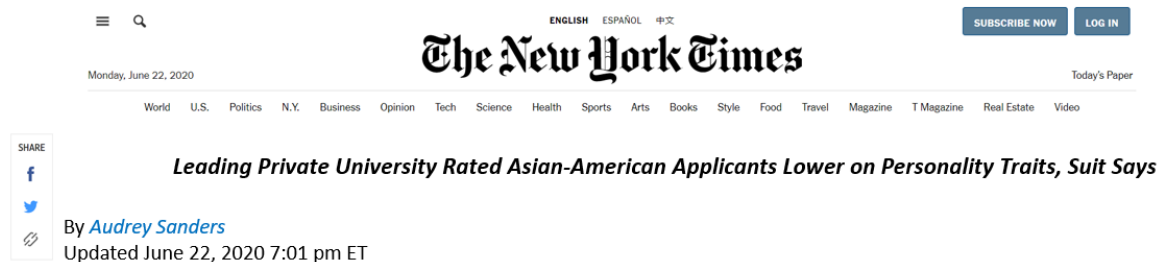
By [Audrey Sanders](#)

Updated June 22, 2020 7:01 pm ET

A leading private university in the United States, admitted only 4.6 percent of its applicants this year. That has led to intense interest in the university's closely guarded admissions process.

The university used an in-house approach for the admissions process, which include both subjective and objective methods. The objective methods include reviewing the applicant's test scores and grades. The subjective methods include analyzing the applicant's traits like "positive personality," likability, courage, kindness and being "widely respected."

The university is now is facing a crisis because their computer algorithms had made a mistake in their objective assessment of student applications. The algorithms had incorrectly used lower test scores for the Asian-American applicants. This resulted in declining the applications of hundreds of Asian-American students of otherwise acceptance-worthy applicants. Now, more than 160,000 student records filed by a group representing Asian-American students in a lawsuit against the university.

Algorithm error in subjective task condition:


Monday, June 22, 2020

World U.S. Politics N.Y. Business Opinion Tech Science Health Sports Arts Books Style Food Travel Magazine T Magazine Real Estate Video

SHARE

Leading Private University Rated Asian-American Applicants Lower on Personality Traits, Suit Says

By [Audrey Sanders](#)

Updated June 22, 2020 7:01 pm ET

A leading private university in the United States, admitted only 4.6 percent of its applicants this year. That has led to intense interest in the university's closely guarded admissions process.

The university used an in-house approach for the admissions process, which include both subjective and objective methods. The objective methods include reviewing the applicant's test scores and grades. The subjective methods include analyzing the applicant's traits like "positive personality," likability, courage, kindness and being "widely respected."

The university is now is facing a crisis because their computer algorithms had made a mistake in their subjective assessment of student applications. The algorithms had incorrectly rated Asian-American applicants lower than others on traits like "positive personality," likability, courage, kindness and being "widely respected." This resulted in declining the applications of hundreds of Asian-American students of otherwise acceptance-worthy applicants. Now, more than 160,000 student records filed by a group representing Asian-American students in a lawsuit against the university.

Study 6 - Stimuli

Human error interactive task condition:

Support The Guardian

Available for everyone, funded by readers

Contribute →

Subscribe →

Search jobs

Sign in

Search

International edition

News

Opinion

Sport

Culture

Lifestyle

More

Fashion

Food

Recipes

Love & sex

Health & fitness

Home & garden

Women

Men

Family

Travel

Money

Fashion

Can D&J Deal with the Crisis Caused by Its Interactive Personal Stylists?



Audrey Sanders

In recent weeks, D&J a leading fashion retailer brand has been facing growing customer complaints because of some problems caused by its interactive personal stylists, a recent introduction to personalize products for customers to not only reflect but also to accentuate their personalities.

Customers say that they were misled by the interactive stylists, as a result of which they bought products that were very expensive and that also did not reflect their personalities and were, in fact, a misfit with their personalities. These customers are now demanding refunds for these products and threatening to sue D&J.

D&J introduced the personal stylists concept in early 2019. Customers who want to use the interactive personal stylists first complete an online form which asks for a personal photograph and details of their height, weight, and personal likes and dislikes of different colors and styles. Then, the D&J personal stylists interact (online) with customers where the customers can see how the products will look on them. The customer is actively involved in the selection of products by the interactive personal stylists. Based on the information provided by the customers, the personal stylist chooses products for the customer which are then shipped to the customer.

The problem arose when the interactive personal stylists had made a mistake in assessing the customers' personal profiles, which resulted in the wrong products being sent to customers.

Human error non-interactive task condition:

Support The Guardian

Available for everyone, funded by readers

Contribute →

Subscribe →

Search jobs

Sign in

Search

International edition

News

Opinion

Sport

Culture

Lifestyle

More

Fashion

Food

Recipes

Love & sex

Health & fitness

Home & garden

Women

Men


Family

Travel

Money

Fashion

Can D&J Deal with the Crisis Caused by Its Personal Stylists?



Audrey Sanders

In recent weeks, D&J a leading fashion retailer brand has been facing growing customer complaints because of some problems caused by its personal stylists, a recent introduction to personalize products for customers to not only reflect but also to accentuate their personalities.

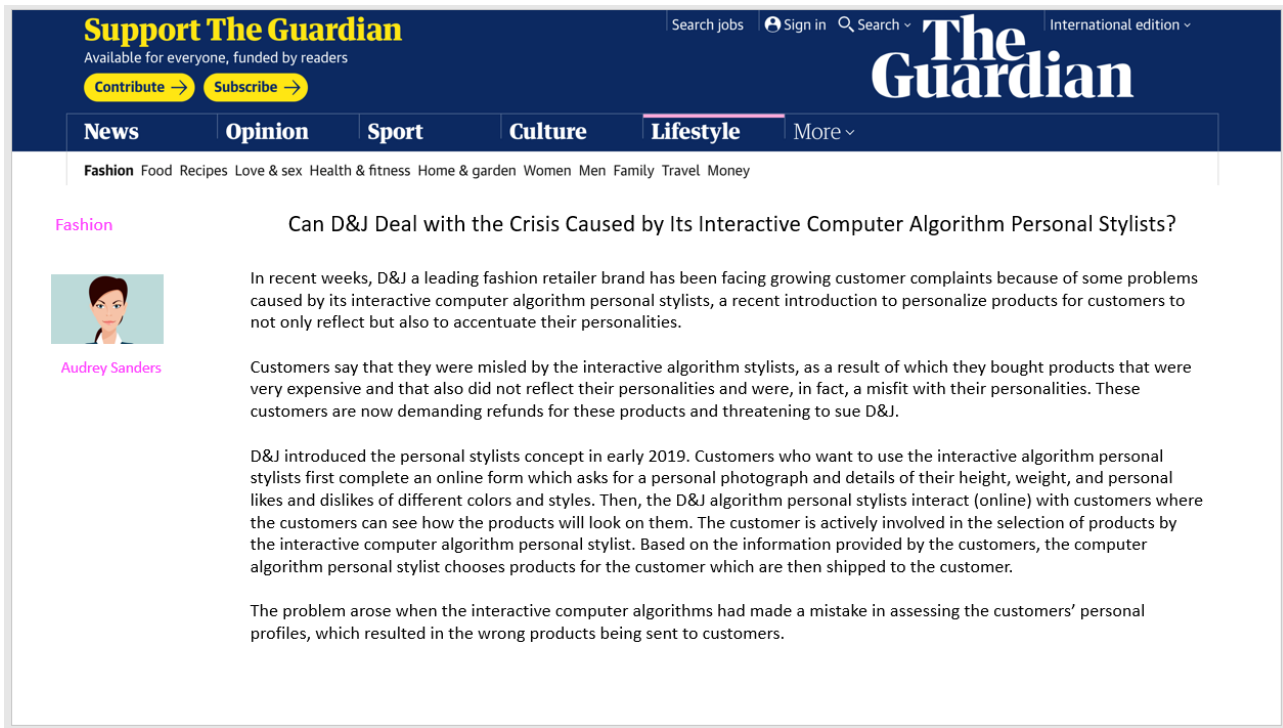
Customers say that they were misled by the stylists, as a result of which they bought products that were very expensive and that also did not reflect their personalities and were, in fact, a misfit with their personalities. These customers are now demanding refunds for these products and threatening to sue D&J.

D&J introduced the personal stylists concept in early 2019. Customers who want to use the personal stylists first complete an online form which asks for a personal photograph and details of their height, weight, and personal likes and dislikes of different colors and styles. Then, it's up to the D&J personal stylists to choose the right products for the customer. The customer is not involved in the selection of the products. Based on the information provided by the customers, the personal stylist chooses products for the customer which are then shipped to the customer.

This particular problem arose when the personal stylists had made a mistake in assessing the customers' personal profiles, which resulted in the wrong products being sent to customers.

Author Accepted Manuscript

Algorithm error interactive task condition:



Support The Guardian
Available for everyone, funded by readers
Contribute → Subscribe →

Search jobs Sign in Search International edition

The Guardian


News Opinion Sport Culture **Lifestyle** More

Fashion Food Recipes Love & sex Health & fitness Home & garden Women Men Family Travel Money

Fashion

Can D&J Deal with the Crisis Caused by Its Interactive Computer Algorithm Personal Stylists?

In recent weeks, D&J a leading fashion retailer brand has been facing growing customer complaints because of some problems caused by its interactive computer algorithm personal stylists, a recent introduction to personalize products for customers to not only reflect but also to accentuate their personalities.

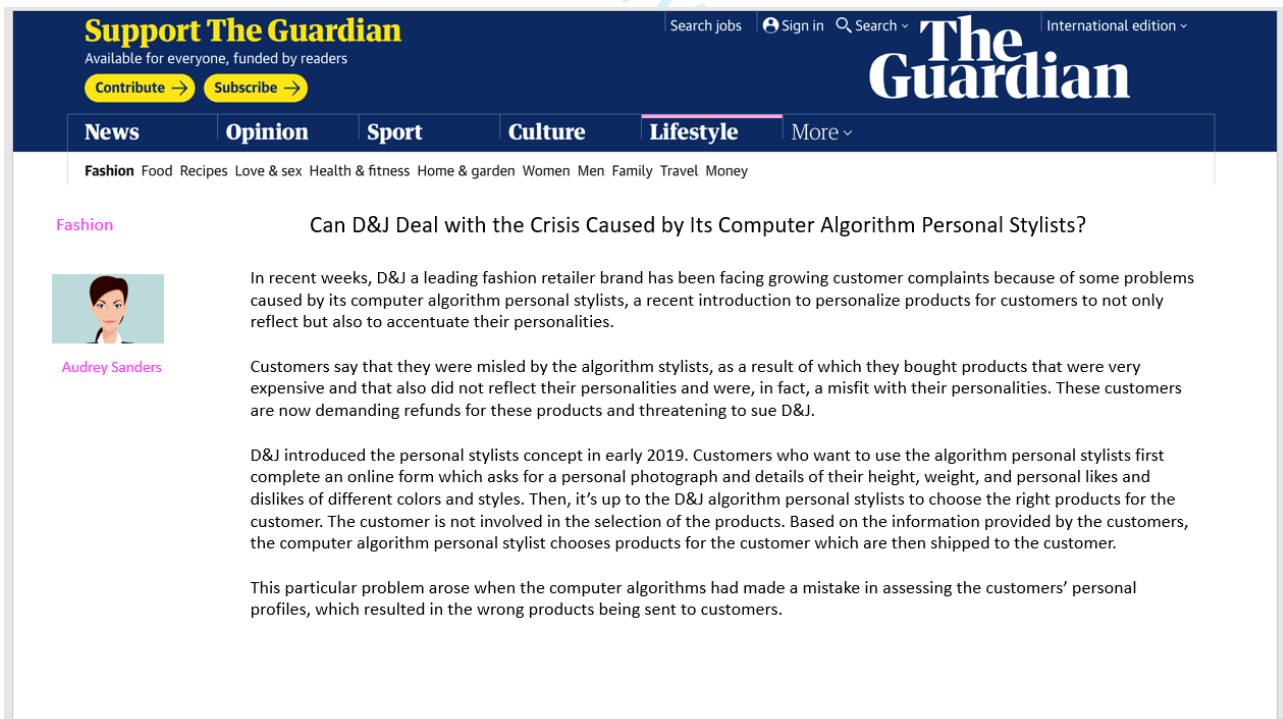

Audrey Sanders

Customers say that they were misled by the interactive algorithm stylists, as a result of which they bought products that were very expensive and that also did not reflect their personalities and were, in fact, a misfit with their personalities. These customers are now demanding refunds for these products and threatening to sue D&J.

D&J introduced the personal stylists concept in early 2019. Customers who want to use the interactive algorithm personal stylists first complete an online form which asks for a personal photograph and details of their height, weight, and personal likes and dislikes of different colors and styles. Then, the D&J algorithm personal stylists interact (online) with customers where the customers can see how the products will look on them. The customer is actively involved in the selection of products by the interactive computer algorithm personal stylist. Based on the information provided by the customers, the computer algorithm personal stylist chooses products for the customer which are then shipped to the customer.

The problem arose when the interactive computer algorithms had made a mistake in assessing the customers' personal profiles, which resulted in the wrong products being sent to customers.

Algorithm error non-interactive task condition:



Support The Guardian
Available for everyone, funded by readers
Contribute → Subscribe →

Search jobs Sign in Search International edition

The Guardian


News Opinion Sport Culture **Lifestyle** More

Fashion Food Recipes Love & sex Health & fitness Home & garden Women Men Family Travel Money

Fashion

Can D&J Deal with the Crisis Caused by Its Computer Algorithm Personal Stylists?

In recent weeks, D&J a leading fashion retailer brand has been facing growing customer complaints because of some problems caused by its computer algorithm personal stylists, a recent introduction to personalize products for customers to not only reflect but also to accentuate their personalities.


Audrey Sanders

Customers say that they were misled by the algorithm stylists, as a result of which they bought products that were very expensive and that also did not reflect their personalities and were, in fact, a misfit with their personalities. These customers are now demanding refunds for these products and threatening to sue D&J.

D&J introduced the personal stylists concept in early 2019. Customers who want to use the algorithm personal stylists first complete an online form which asks for a personal photograph and details of their height, weight, and personal likes and dislikes of different colors and styles. Then, it's up to the D&J algorithm personal stylists to choose the right products for the customer. The customer is not involved in the selection of the products. Based on the information provided by the customers, the computer algorithm personal stylist chooses products for the customer which are then shipped to the customer.

This particular problem arose when the computer algorithms had made a mistake in assessing the customers' personal profiles, which resulted in the wrong products being sent to customers.

Study M1 – Stimuli**Algorithm error condition****Fiat Chrysler recalls 4.8 million US cars**

Fiat Chrysler is recalling 4.8 million US vehicles over a defect that could prevent drivers from turning off cruise control. It warned owners not to use the function until they get software upgrades.

Most of the vehicles being recalled cover models built between 2014-2018. The spokesman of Fiat Chrysler noted that no injuries or crashes were related because of this product recall.

Cruise control system was developed by a computer algorithm at Fiat Chrysler. The computer algorithm at Fiat Chrysler had made a mistake resulting in a defect in the cruise control system causing a safety hazard.

Human error condition**Fiat Chrysler recalls 4.8 million US cars**

Fiat Chrysler is recalling 4.8 million US vehicles over a defect that could prevent drivers from turning off cruise control. It warned owners not to use the function until they get software upgrades.

Most of the vehicles being recalled cover models built between 2014-2018. The spokesman of Fiat Chrysler noted that no injuries or crashes were related because of this product recall.

Cruise control system was developed by employees of Fiat Chrysler. Fiat Chrysler employees had made a mistake resulting in a defect in the cruise control system causing a safety hazard.

Human supervision condition**Fiat Chrysler will have more managerial supervision over errors**

Following the recall of 4.8 million US vehicles over a defect that could prevent drivers from turning off cruise control, Fiat Chrysler issued a statement that "The company will inform the owners of defective cars with a registered letter. Moreover, we note that, going forward, there will be increased managerial supervision in the Fiat Chrysler manufacturing processes. We anticipate that these additional checks and balances with additional human supervision will ensure superior product quality and fewer defects."

Technological supervision condition**Fiat Chrysler will have more technological supervision over errors**

Following the recall of 4.8 million US vehicles over a defect that could prevent drivers from turning off cruise control, Fiat Chrysler issued a statement that "The company will inform the owners of defective cars with a registered letter. Moreover, we note that, going forward, there will be increased technological supervision in the Fiat Chrysler manufacturing processes. We anticipate that these additional checks and balances with additional technological supervision will ensure superior product quality and fewer defects."