Computers in Human Behavior 122 (2021) 106859

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Threat of racial and economic inequality increases preference for algorithm decision-making

Yochanan E. Bigman [a,b,*], Kai Chi Yam [c], Déborah Marciano [d], Scott J. Reynolds [e], Kurt Gray [a]

[a] University of North Carolina at Chapel Hill, USA
[b] Yale University, USA
[c] National University of Singapore, Singapore
[d] University of California, Berkeley, USA
[e] University of Washington, USA

ARTICLE INFO

Keywords:
Algorithm aversion
COVID-19
Inequality threat
Health disparities

ABSTRACT

Artificial intelligence (AI) algorithms hold promise to reduce inequalities across race and socioeconomic status. One of the most important domains of racial and economic inequalities is medical outcomes; Black and low-income people are more likely to die from many diseases. Algorithms can help reduce these inequalities because they are less likely than human doctors to make biased decisions. Unfortunately, people are generally averse to algorithms making important moral decisions—including in medicine—undermining the adoption of AI in healthcare. Here we use the COVID-19 pandemic to examine whether the threat of racial and economic inequality increases the preference for algorithm decision-making. Four studies (N = 2819) conducted in the United States and Singapore show that emphasizing inequality in medical outcomes increases the preference for algorithm decision-making for triage decisions. These studies suggest that one way to increase the acceptance of AI in healthcare is to emphasize the threat of inequality and its negative outcomes associated with human decision-making.

1. Introduction

Large racial and economic inequalities exist both across and within nations. In the United States, Black men are 2.5 times more likely to die of police violence than White men (Edwards et al., 2019). In the US, 31.4% of national wealth is owned by the top 1% of the population (Board of Governors of the Federal Reserve System, 2021). In Singapore, the median salary for Singaporean employees is over six times that of migrant workers (Geddi et al., 2020). These racial and economic disparities are also reflected in medical outcomes. In 2020, in the United States the life expectancy at birth of non-Hispanic Black people was 6 years lower than that of non-Hispanic White people, and 8 years lower than that of Hispanic people (Arias et al., 2021). Similarly, Black people in the United States are 30% more likely to die from a heart disease than non-Hispanic White people (Murphey et al., 2021). One way to somewhat reduce inequality in medicine is to rely on artificial intelligence (AI)-powered algorithms to make medical decisions (Hao, 2020), and here we use the context of COVID-19 to examine whether the threat of racial and economic inequality might increase preference for algorithm decision-making.

The COVID-19 pandemic has amplified existing health disparities. In the middle of the first wave of infection, the COVID-19 death rate in the United States was 61.6 deaths (per 100,000 people) for Black Americans and 26.2 for White Americans (APM Research Lab, 2020). In Singapore, more than 90% of infections comprising migrant workers who make less than SGD$10,000 annually (~USD$7500; Palma, 2020). Similar disparities exist in other countries, with the poor and minorities being most affected by the virus (Bhala et al., 2020). Health disparities—both in general and in COVID-19—are tied to habitat density, pre-pandemic healthcare, healthcare access, and the ability to work remotely (Abuelgasim et al., 2020; Bibbins-Domingo, 2020; Braveman et al., 2011; Yancy, 2020), but may also stem, at least in a small part, from bias – "prejudice in favor of or against one thing, person, or group compared with another, usually in a way considered to be unfair" (Oxford University Press, 2020) – in human decision-making. For example, empirical studies have demonstrated that doctors sometimes under-screen, under-diagnose, and under-treat members of some minority groups (Alsan et al., 2019; Hoffman et al., 2016).

* Corresponding author. Department of Psychology, Yale University, New Haven, CT, USA.
E-mail address: yochanan.bigman@yale.edu (Y.E. Bigman).

https://doi.org/10.1016/j.chb.2021.106859
Received 15 February 2021; Received in revised form 8 April 2021; Accepted 4 May 2021
Available online 6 May 2021

Many have suggested that using AI can help reduce bias in human decision-making (Houser, 2019; Munoz et al., 2016). While the idea of AI being in charge of patients might appear futuristic, AI has been used for a few years already in emergency rooms around the world, for example in London (Crouch, 2019) and Cleveland (Gauher & Uz, 2016). Of course, in both medical contexts and non-medical contexts where AI is used (e.g., the military, finance), it is always humans who have ultimate responsibility for making decisions. But people may start to prefer that AI systems are used more in these decision-making contexts, and we use "algorithm decision-making" as a short-hand for this idea.

Recent studies have examined the role of AI in the treatment and management of COVID-19, revealing that algorithms can improve the work of healthcare workers and the outcomes of patients, ranging from more accurate diagnoses and more efficient monitoring (Tayarani-N., 2020). Healthcare workers generally appreciate the efficiency of algorithms (Ardon & Schmidt, 2020; Laï et al., 2020; Polesie et al., 2020), although they do worry about AI's lack of empathy and its low ability to communicate with patients (Blease et al., 2019; Doraiswamy et al., 2020).

Hospitals currently rely on AIs for a wide range of needs, including informing whether an individual should get tested for COVID-19 (Gerretsen, 2020; Meah, 2020; Parrock, 2020; Vanian, 2020). A recent study also found that for triage decisions,[1]—which we define broadly as the identification of patients who are most at risk and in need of appropriate treatments,—AI can accurately predict the risk of COVID-19 patients developing critical illness (Liang et al., 2020). In other words, early AI-informed triage is already a reality in several hospitals (Hao, 2020; Vanian, 2020) and will only grow in popularity. Interestingly, the use of AIs can reduce the need for "classic" life and death triage decisions which are characterized by a lack of resources and time pressure.

Using AI can increase the efficiency of medical diagnoses, and by doing so prevent using scarce resources such as ventilators on patients who do not have a medical need for them. This can allow hospitals to conserve resources for patients in dire need of them. Of course, algorithms are not free of bias (Angwin et al., 2016; Dastin, 2018; Lambrecht & Tucker, 2019). For example, the algorithms used to identify patients for special medical treatment systemically underdiagnosed Black people (Obermeyer et al., 2019). However, algorithms are generally *less* biased than humans, and more easily corrected (Mullainathan, 2019; Shea et al., 2020). Even simple algorithms can outperform humans in decision-making tasks by consistently following decision rules rather than relying on intuition as humans sometimes do (Dawes, 1979). However, there is a key barrier to the adoption of AI in medicine: research in psychology reveals that people are generally averse to algorithms making decisions (Dietvorst et al., 2015), especially in medical and moral contexts (Bigman & Gray, 2018; Castelo et al., 2019; Longoni et al., 2019), where empathy (Bigman & Gray, 2018) and viewing the patient as a unique person (Longoni et al., 2019) are seen as important.

Reducing the aversion to algorithm decision-making is especially important during pandemics such as COVID-19, because of the potential for algorithms to improve efficiency and reduce biases (Shea et al., 2020). For example, in the US, human doctors sometimes are less likely to recommend potentially lifesaving screening procedures to Black people (Alsan et al., 2019), perceive Black people as experiencing less pain (Hoffman et al., 2016), and prescribe less pain medication to Black people than their non-Black peers (Morrison et al., 2000). Specific to COVID-19, such biases can impact mortality rates by influencing the allocation of scarce ventilators, ICU beds, and other life-saving resources. By reducing these biases, AIs can potentially improve health outcomes.

How can we reduce the aversion from algorithm decision-making? We draw from social cognitive theory (Fiske & Taylor, 1991; Jones, 1991; Reynolds, 2006) to propose one potential way: highlighting a threat of inequality. Social cognitive theory (Fiske & Taylor, 1991) argues that the extent to which an individual pays attention to information depends on three factors: vividness, salience, and accessibility. Vividness refers to how interesting, provoking, and proximate the information is; salience refers to how noticeable or important the information is; accessibility refers to the extent to which an individual has personal traits or experiences that facilitate a connection to the information. Behavioral ethics scholars have relied on this theory to explain moral awareness, or why an individual would identify an issue as a moral one (Reynolds & Miller, 2015). As Jones (1991) argues, when an issue's characteristics are vivid and salient to the moral domain (e.g., proximate physical harm), an individual is more likely to identify the issue as a moral issue. Reynolds (2006) further demonstrates that individual decision-making frameworks can increase the accessibility of the issue, thus allowing some people to see moral issues where others do not. Importantly, moral awareness is critical because it constitutes the first step in the moral decision-making process, which culminates in moral behavior (Rest, 1986).

We argue that threats of inequality are vivid and salient information that lead people to view some medical decisions as moral issues. While past research has demonstrated that moral concerns can make people averse to algorithm decision-making because algorithms are seen as devoid of emotion (Bigman & Gray, 2018; Young & Monroe, 2019), we suggest that this "cold impartiality" may be a positive factor when inequality is salient. Highlighting inequality in human-led decision--making should motivate people towards an alternative that is perceived as less biased: algorithms.

Hypothesis 1. Threats of racial and economic inequality will increase people's preference for algorithm decision-making.

Social cognitive theory suggests that while threats of inequality might increase the overall preference for algorithm decision-making, the effect may vary at the individual level (Fiske & Taylor, 1991; Reynolds, 2008). Accessibility—an individual's capacity to connect with information based on existing beliefs, traits, and experiences—shapes the extent to which different individuals pay differing amounts of attention to a piece of information. Although the threat of inequality may reduce the aversion to algorithm decision-making in all people, members of the discriminated-against group may show an even stronger effect. For example, one reason Black Americans are more supportive than White Americans of affirmative action is that they believe it offers Black Americans a fairer opportunity in college admissions (Mangum, 2008). Therefore, we predict that while threats of inequality will increase all people's preference for algorithm decision-making, the extent to which the inequality is personally relevant will moderate this effect.

Hypothesis 2. Personal relevance moderates the effect of threat of inequality on preference for algorithm decision-making, such that the increase in preference for algorithms is stronger for members of the disadvantaged group.

How exactly do threats of inequality increase the preference for algorithm decision-making? We suggest that one possible mechanism is that such threats weaken the perceived authority of doctors and healthcare workers. The principle of authority is a core moral value and many people view submitting to legitimate authority as morally obligatory (Graham et al., 2011). Doctors are generally perceived as authority figures, and many people hold them in very high regard (Brase & Richmond, 2004), sometimes even elevating them to a "godlike" status (Goranson et al., 2020). During the COVID-19 pandemic, many in society have referred to doctors as "heroes" (Bauchner & Easley, 2020).

In the case of outcome inequalities, however, tensions emerge because people might see doctors as partially responsible, and that

[1] We acknowledge that in the medical context, triage decisions often refer to decisions where a lack of resources require decisions that result in determining who could be saved or treated. We took a broader definition because we used the term "triage" in our empirical studies where participants were all laymen and not medical professionals.

might violate another core value: the principle of justice (Rawls, 1971). We suggest that because the threat of inequality—a potential violation of the principle of justice—constitutes a vivid, salient, and immediate concern, the individual will give its violation greater attention than a less vivid and less salient obligation to follow authority. In other words, we suggest that people exposed to inequality associated with human doctor's decision-making will have a decreased sense of the authority of doctors. Thus, threats of inequality will weaken the perceived authority of doctors, which will in turn increase the preference for algorithm decision-making.

Hypothesis 3. Authority of doctors mediates the interactive effect of threat of inequality and personal relevance on algorithm aversion.

This research makes several contributions. First, it contributes to the growing literature on the effects of exposure to inequality (McCall et al., 2017; Sands, 2017), revealing a consequence of exposure to health inequalities, and how exposure to this inequality might affect different groups differently. Second, it contributes to the literature on algorithm aversion (Bigman & Gray, 2018; Dietvorst et al., 2015; Longoni et al., 2019) by revealing ways to reduce the aversion. Third, it contributes to our knowledge of the social effects of the belief in the authority of doctors by showing that disparities in health outcomes reduce the perceived authority of doctors. In doing so, we demonstrate an important outcome of emphasizing health disparities.

2. Current research

We tested our hypotheses in four studies and across two cultural contexts of inequality: race in the United States (Studies 1, 3, and 4) and SES in Singapore (Study 2). In all studies we asked participants to imagine that they were experiencing COVID-19 symptoms and needed to decide whether to go to a hospital where either a human doctor or algorithm makes triage decisions. We manipulated whether participants read (real) information about race (Studies 1, 3, and 4) or SES (Study 2) disparities in COVID-19 health outcomes or not. In Study 3 we further examined the mediating role of the perceived authority of doctors. In Study 4 we used a binary choice measure, rather than a continuous preference measure. The studies were approved by the University of North Carolina at Chapel Hill IRB. Demographics of each study appear in Table 1. Data, supplemental materials, and full study materials are available online at https://osf.io/mhjdy/?view_only=7d1442e70df9426594c1daee10b73e2f.

3. Study 1: race disparities in the US

At the time we conducted this study (see Table 1) the COVID-19 mortality rate of Black Americans in the United States was more than twice that of White Americans (APM Research Lab, 2020). In Chicago, for example, Black Americans account for only 33% of the population, but 72% of the deaths attributed to COVID-19 (Eligon et al., 2020). In Study 1 we tested whether the threat of inequality—manipulated by informing participants about COVID-19 racial health disparities—increased people's preference for algorithm decision-making and whether personal relevance (the race of the participant) moderates this effect.

3.1. Method

3.1.1. Participants

We recruited 1379 participants[2] through Amazon's Mturk, using Turkprime (Litman et al., 2017). We aimed for a half-half ratio of Black and White participants in our samples (with specified samples). After excluding participants who failed the attention check, manipulation comprehension check, or did not self-identify as White or Black, we ended up with 1060 participants (481 male, 572 female, 7 other; 535 White and 525 Black; Age: $M = 37.41$, $SD = 12.55$). Results remain unchanged when the analyses were performed on the full sample. Participants were paid between $0.50 and $1. A post-hoc power analysis (using G*power 3.1, Faul et al., 2007) revealed that this sample size had an achieved power of 1 (for an alpha of .05) to detect the main effect for condition, and an achieved power of .995 to detect the interaction.

3.1.2. Threat of inequality manipulation

After consenting and completing the first attention check in which they were asked what day of the week it was yesterday and what they had for breakfast that day, participants were randomly assigned to either a control condition or threat of inequality condition. In both conditions participants read the following text:

As COVID-19 spreads, hospitals are running low on crucial life-saving equipment, such as ventilators. When there are not enough resources to give all patients the care they need, "triage" is used to determine who gets priority in medical care. Triage decisions can be made by either human doctors or specialized algorithms.

In the threat of inequality condition participants then also read the following text (in the control condition participants were directed to the measures), based on real data on COVID-19 mortality rates at the time of the study (Eligon et al., 2020) and on research suggesting a bias in medical care towards Black people (Alsan et al., 2019; Hoffman et al., 2016):

In America, there are large racial disparities in who dies from COVID19. In Chicago, for example, African Americans account for only 33% of the population, but 72% of COVID19-related deaths. It is currently unclear what causes this disparity, but there is some evidence suggesting that generally white doctors tend to under-diagnose African American patients.

3.1.3. Preference for algorithm decision-making

All participants then read the following question:

Imagine you are feeling severe shortness of breath and need to go to a hospital.There are two nearby hospitals. You know that both hospitals are running low on supplies and need to prioritize patients. In one hospital a human doctor makes triage decisions; in the second hospital an AI-based algorithm makes triage decisions. To which hospital would you go?

Participants answered the question on a 1 ("Definitely the hospital where the human doctor makes triage decisions") to 5 ("Definitely the hospital where an AI-based algorithm makes triage decisions") scale (see Longoni et al., 2019, for a similar measurement).[3]

Participants then completed other measures (see supplemental materials) and the manipulation comprehension check, in which they were asked if the scenario they read mentioned race disparities in COVID-19 mortality rate (yes/no). Finally, participants provided demographic information.

3.2. Results

Descriptive statistics and correlations are presented in Table 2.

[2] Participants in Study 1 were recruited in 3 different samples. All samples had an identical manipulation and dependent variable, but included different exploratory variables. See sample details and full study materials for each sample in the supplemental materials.

[3] We note that in Sample A we used a 1 to 7 rather than a 1 to 5 scale. When combining the data of Sample A with Samples B and C we transformed the responses from Sample A to a 1–5 scale, using the following transformation: Xnew = (Xold-1)*4/6 + 1, such that 1 on the 1–7 scale would be 1 on the 1–5 scale, 7 would be transformed to 5, and 4, the midpoint in the 1–7 scale, to 3, the midpoint in the 1–5 scale. Analyzing the samples separately or combined yield essentially identical results.

Table 1
Sample details for Studies 1–4.

	Study 1 Sample A	Study 1 Sample B	Study 1 Sample C	Study 2	Study 3	Study 4
Initial N	408	486	485	662	601	1005
Final N	295	377	388	416	483	860
Age (SD)	35.23 (11.91)	39.01 (12.81)	37.51 (12.57)	21.95 (1.58)	38.29 (11.71)	32.05 (10.54)
Gender	129 Male 165 Female 1 Other	167 Male 206 Female 4 Other	185 Male 201 Female 2 Other	153 Male 257 Female 6 Other	214 Male 266 Female 3 Other	432 Male 419 Female 9 Other
Race	155 White 140 Black	183 White 194 Black	197 White 191 Black		231 White 252 Black	446 White 208 Black 190 Latino 16 Indigenous
Date (2020)	May 4–6	May 14–15	May 27–28	June 11	June 24–26	November 9

Table 2
Descriptive statistics and correlations for study variables (study 1).

Variable	Mean	SD	1	2	3
1. Threat of inequality	.48	.50	(- -)		
2. Personal relevance	.50	.50	.07*	(- -)	
3. Preference for algorithm triage	2.59	1.62	.27*	.13*	(- -)

Notes.
Threat of inequality: 0 = control condition; 1 = threat of inequality.
Personal relevance: 0 = White participants; 1 = Black participants.
*p < .05.

A 2 (condition: inequality threat, control) x 2 (personal relevance: Black, White) ANOVA revealed a main effect for condition, $F(1, 1056) = 78.45$, $p < .001$, $\eta_p^2 = 0.07$, such that participants reported a stronger preference for algorithm decision-making in the threat of inequality condition ($M = 2.77$, $SD = 1.43$) than in the control condition ($M = 2.04$, $SD = 1.20$), supporting Hypothesis 1.

We also found a main effect for personal relevance, $F(1, 1056) = 17.05$, $p < .001$, $\eta_p^2 = 0.02$, such that across conditions Black participants preferred algorithm decision-making ($M = 2.58$, $SD = 1.47$) more than White participants ($M = 2.21$, $SD = 1.22$). Finally, the personal relevance × condition interaction was also significant, $F(1, 1056) = 20.88$, $p < .001$, $\eta_p^2 = 0.02$, such that while both Black and White participants preferred algorithm decision-making more in the threat of inequality condition, this effect was stronger for Black participants (threat of inequality: $M = 3.09$, $SD = 1.45$; control: $M = 2.02$, $SD = 1.19$; $F(1,1056) = 89.67$, $p < .001$, $\eta_p^2 = 0.08$) than White participants (threat of inequality: $M = 2.40$, $SD = 1.13$; control: $M = 2.06$, $SD = 1.12$; $F(1, 1056) = 9.24$, $p < .001$, $\eta_p^2 = 0.01$), supporting Hypothesis 2 (Fig. 1).

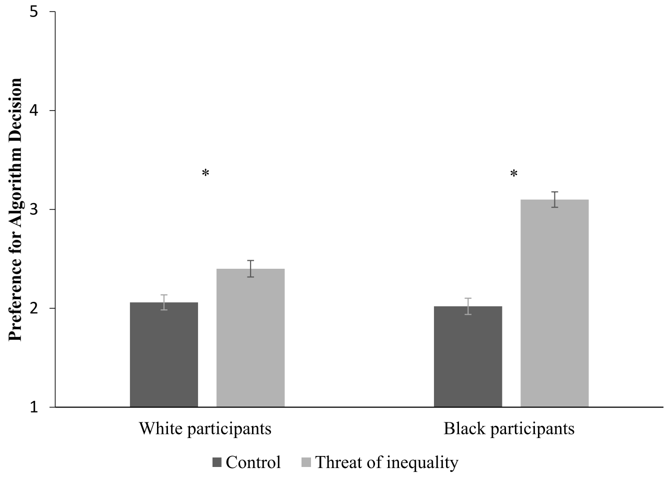

Fig. 1. The interactive effect of threat of inequality and personal relevance on preference for algorithm decision-making (Study 1). Error bars reflect standard errors.

3.3. *Discussion*

The results from Study 1 support Hypotheses 1 and 2: threat of inequality increased preference for algorithm decision-making, especially for those who could be personally disadvantaged by the inequality. Importantly, we note that for Black participants in the threat of inequality condition, the average rating was around the mid-point (3 on the 1 to 5 scale), suggesting that these participants overcame the algorithm aversion reported in previous research (e.g., Bigman & Gray, 2018; Dietvorst et al., 2015). However, one limitation of Study 1 is that it focuses on a threat of inequality in a specific cultural context—racial health disparities in the United States (although we did collect most of this data before the unrest in the United States following the murder of George Floyd on May 25, 2020). In Study 2 we address this limitation by focusing on a different cultural context—SES health disparities in Singapore.

4. Study 2: class disparities in Singapore

In Study 2, we operationalized the threat of inequality through social class. As with race, COVID-19 has disproportionally affected the poor. For example, in the UK those infected with COVID-19 who live in the poorest 10% areas of the country have been twice as likely to die as those living in the wealthiest 10% areas of the country (Devlin & Barr, 2020). In this light, we tested our hypotheses in Singapore, a country where COVID-19 has infected more than 40,000 people as of June 18, 2020, with more than 90% of them being low-paid migrant workers (Palma, 2020). This context presents an ideal test case for our hypotheses with high realism. Another limitation of Study 1 is that we informed participants about possible bias in doctors' decision-making, which might have created demand characteristics. In Study 2 we minimized these possible demand characteristics by testing our hypotheses without disclosing the source of the health disparities.

4.1. *Method*

4.1.1. *Participants*

We recruited 662 undergraduates from a university in Singapore. Excluding participants who failed the manipulation comprehension check, we ended up with 416 participants (153 males, 257 females, and 6 other/preferred not to disclose; Age: $M = 22.04$, $SD = 1.58$). Results remain unchanged when the analyses were performed on the full sample. Participants were paid SGD$5 (~USD$3.60) in exchange for their participation. A post-hoc power analysis (using G*power 3.1, Faul et al., 2007) revealed that this sample size had an achieved power of .985 (for an alpha of .05) to detect the main effect for condition, and an achieved power of .999 to detect the interaction. The study was conducted in English, the language of instruction in the university from which our sample was taken. In addition, all participants were fluent in English.

4.1.2. SES

Participants first completed the MacArthur subjective SES scale (Adler et al., 2000), which is an established measure of SES widely used in social psychology (e.g., Dubois et al., 2015; Piff et al., 2012). They were asked to position themselves on a ladder representing where people in Singapore stand in relation to education, wealth, and respectful jobs (i.e., the definition of SES), on a 1 ("I am the worst off") to 10 ("I am the best off").

4.1.3. Threat of inequality manipulation

We then randomly assigned participants to either a threat of inequality condition or a control condition. The control condition was identical to that in Study 1 in which participants read about the scarcity of medical resources needed for COVID-19 treatment. In the threat of inequality condition, participants also read the following text (in the control condition participants were directed to the measures), which provided real data about Singapore at the time of the study:

In Singapore, there are large class disparities in who are infected with COVID-19. Out of the 35,000 confirmed cases, over 95% are among Singapore's lowest-paid foreign workers. They live in dormitories located and almost hidden from view on the outskirts of the city, with poor and overcrowded living conditions. They sleep on bunk beds, 12 to 20 packed into one room, poorly ventilated by small fans and communal toilets and showering facilities shared by hundreds of men on each floor. Typically, they earn less than SGD 1000 a month, while the median monthly income for Singaporeans is SGD 3227.

4.1.4. Preference for algorithm decision-making

We measured preference for algorithm decision-making as in Study 1. Participants then completed other scales (see supplemental materials) and, as a manipulation comprehension check, were asked whether the scenario they read mentioned class disparities in COVID-19 (yes/no). Finally, participants provided demographic information.

4.2. Results

Descriptive statistics and correlations are presented in Table 3. The mean SES in our sample was 6.22 ($SD = 1.43$), suggesting that participants perceived themselves as slightly higher than average SES; this makes sense since participants were undergraduates in an elite university in Singapore. This analysis also provides a conservative test for our hypotheses, as our participants were on a higher-than-average SES and therefore not strongly disadvantaged.

A *t*-test revealed that participants reported a stronger preference for algorithm decision-making in the threat of inequality condition ($M = 2.41$, $SD = 1.34$) than in the control condition ($M = 1.96$, $SD = 0.98$), $t(414) = 3.92$, $p < .001$, Cohen's $d = 0.38$, supporting Hypothesis 1.

To test the hypothesized interaction between SES and threat of inequality, we conducted a step-wise OLS regression predicting preference for algorithm decision-making. In step 1, we entered SES and threat of inequality as the independent variables. In step 2, we entered the interaction term. As expected, the interaction term was significant ($\beta = -0.21, p < .001$) and resulted in a significant change in R^2 (adjusted $R^2 = 0.11$, $\Delta R^2 = 0.04$, $F_{change}(1, 412) = 20.26, p < .001$). Follow-up simple slope tests suggest that there is a significant and positive association between threat of inequality and preference for algorithm decision-making for those with low subjective SES (-1SD; $t = 6.10$, $p < .001$), but not for those with high subjective SES (+1SD; $t = -0.29$, $p = .77$; Fig. 2), supporting Hypothesis 2.

Table 3
Descriptive statistics and correlations for study variables (study 2).

Variable	Mean	SD	1	2	3
1. Threat of inequality	.49	.50	(- -)		
2. Personal relevance (SES)	6.22	1.43	.01	(- -)	
3. Preference for algorithm	2.18	1.18	.19*	-.20*	(- -)

Notes.
Threat of inequality: 0 = control condition; 1 = threat of inequality.
*$p < .05$.

4.2.1. Discussion

The results of Study 2 provide further support for our hypotheses in the context of SES health disparities in Singapore. Threat of inequality decreased participants' aversion from algorithm decision-making, especially for those low in subjective SES. These results were obtained even though our participants were university students who were not members of the group the manipulation described as being disadvantaged (i.e., migrant workers). These findings support our hypotheses in a second cultural context, suggesting their generalizability. Furthermore, Study 2 shows that the threat of inequality increases preference for algorithm decision-making even when the biases of human doctors are not explicitly mentioned.

5. Study 3: perceived authority of doctors as a mediator

Studies 1–2 did not explore the psychological mechanism underlying the effect. In Study 3 we test whether the threat of inequality changes people's perceptions of doctors' authority as a mediating mechanism for these effects.

5.1. Method

5.1.1. Participants

We recruited 601 participants through Amazon's Mechanical Turk, using Turkprime (Litman et al., 2017). We aimed for a half-half ratio of Black and White participants in our samples (with specified samples). As specified in the pre-registration[4] (https://aspredicted.org/63pu8.pdf), we excluded participants who did not self-report as either Black or White or that failed the attention and manipulation comprehension checks. When the analyses were performed on the full sample, the effect of condition did not reach significance ($p = .086$) but was in the predicted and pre-registered direction. Our final sample was 483 participants (214 male, 266 female, 3 other; 231 White and 252 Black; Age: $M = 38.29$, *SD*

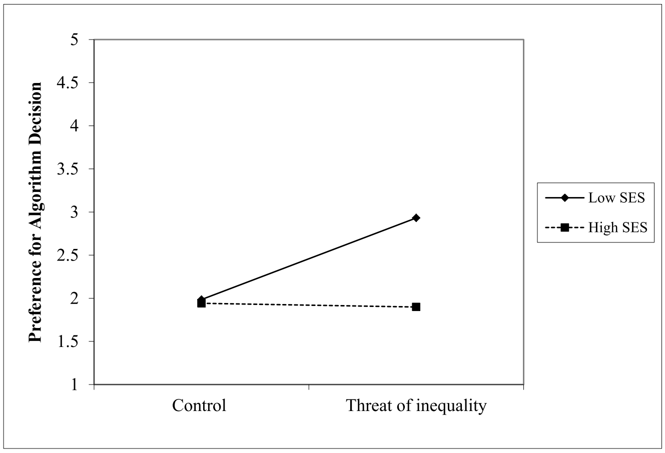

Fig. 2. The interactive effect of inequality threat and SES on preference for algorithm decision-making (Study 2).

[4] As specified in the pre-registration, we measured trust in physicians (Thom et al., 1999) as another possible mediator. The results of that measure were not significant and we do not discuss them further here. See supplemental materials for a full description of the scale.

= 11.71), see Table 1. Participants were paid $1 as compensation. A post-hoc power analysis (using G*power 3.1, Faul et al., 2007) revealed that this sample size had an achieved power of .868 (for an alpha of .05) to detect the main effect for condition, and an achieved power of .382 to detect the interaction.

5.1.2. *Threat of inequality manipulation*

The threat of inequality manipulation was identical to that of Study 1.

5.1.3. *Authority*

After the threat of inequality manipulation participants answered three items measuring how they perceive the authority of doctors, modified from the Moral Foundations Questionnaire (Graham et al., 2011) on a 1 (Strongly disagree) to 5 (Strongly agree) scale: "Respect for the authority of doctors is something all kids should learn," "I am completely comfortable submitting to the authority of doctors," and "As a patient, if I disagree with a doctor's orders, I should still obey anyway because that is my duty," Cronbach's $\alpha = .73$.

5.1.4. *Preference for algorithm decision-making*

We measured preference for algorithm decision-making as in Study 1.

5.1.5. *Attention checks*

In addition to the manipulation comprehension question used in Study 1, we asked participants to explain their choice (excluding participants who provided irrelevant explanations such as "good survey"). We further excluded participants who used auto-completion answering scales by measuring the number of clicks on some screens (using Qualtrics' Timing option). Finally, participants provided demographic information.

5.2. *Results*

Descriptive statistics and correlations are presented in Table 4.

5.2.1. *Preference for algorithm decision-making*

A 2 (condition: threat of inequality, control) x 2 (personal relevance: Black, White) ANOVA revealed a main effect for condition, $F(1, 479) = 9.52$, $p = .002$, $\eta_p^2 = 0.02$, such that participants reported a stronger preference for algorithm decision-making in the threat of inequality condition ($M = 2.39$, $SD = 1.32$) than in the control condition ($M = 2.03$, $SD = 1.22$), supporting Hypothesis 1.

Personal relevance was also significant, $F(1, 479) = 4.49$, $p < .001$, $\eta_p^2 = 0.01$, such that across conditions, Black participants preferred algorithm decision-making ($M = 2.36$, $SD = 1.37$) more than White participants ($M = 2.09$, $SD = 1.16$). The personal relevance × condition interaction was not significant, $F(1, 479) = 2.76$, $p = .097$, which does not support Hypothesis 2. Although the interaction was not significant, it did trend in a similar direction to that of Studies 1–2. For White participants, there was no significant difference between the threat of inequality condition and the control condition (*M* = the control condition ($M = 2.00$, $SD = 1.17$) and 2.17, $SD = 1.15$; $F(1,479) = 0.97$, $p = .325$); For Black participants, there was a significant difference, such that in the threat condition participants showed a greater preference for algorithm decision-making ($M = 2.60$, $SD = 1.43$) than in the control condition ($M = 2.05$, $SD = 1.27$; $F(1,479) = 11.77$, $p < .001$, $\eta_p{}^2 = .024$), see Fig. 3.

Table 4
Descriptive statistics and correlations for study variables (study 3).

Variable	Mean	SD	1	2	3	4
1. Threat of inequality	.55	.50	(- -)			
2. Personal relevance	.52	.50	.02	(- -)		
3. Authority	3.23	.93	-.05	-.06	(- -)	
4. Preference for algorithm	2.23	1.29	.14*	.10*	-.16*	(- -)

Notes.
Threat of inequality: 0 = control condition; 1 = threat of inequality.
Personal relevance: 0 = White participants; 1 = Black participants.
*p < .05.

5.2.2. *Authority*

A 2 (personal relevance: Black, White) x 2 (condition: threat of inequality, control) ANOVA revealed a significant personal relevance × condition interaction, $F(1, 479) = 5.24$, $p = .023$, $\eta_p^2 = .01$, such that while Black participants reported a lower belief in doctors' authority in the threat of inequality condition ($M = 3.05$, $SD = 0.94$) than the control condition ($M = 3.34$, $SD = 0.96$), $F(1, 479) = 6.04$, $p = .014$, $\eta_p^2 = .01$, there was no significant difference for White participants between the threat of inequality condition ($M = 3.33$, $SD = 0.84$) and the control condition ($M = 3.23$, $SD = 0.97$), $p = .415$. See Fig. 4.

5.2.3. *Moderated mediation by authority*

To examine the first-stage moderated mediation (Edwards & Lambert, 2007), we entered threat of inequality as the independent variable, authority as the mediator, personal relevance as the first-stage moderator, and preference for algorithm decision-making as the dependent variable into a bootstrapping moderated mediation analysis (Model 7, 5000 iterations; Preacher & Hayes, 2008). Results revealed that the indirect effect of threat of inequality on preference for algorithm decision-making, via authority, was significant for Black participants (coefficient = 0.06, $SE = 0.03$, 95% CI [0.01, 0.13], but not for White participants (coefficient = −0.02, $SE = 0.03$, 95% CI [-0.08, 0.02]. Finally, the difference between these two indirect effects was also significant (index of moderated mediation = 0.08, $SE = 0.04$, 95% CI [0.01, .19], supporting Hypothesis 3.

5.3. *Discussion*

The results of Study 3 replicate our previous finding that threat of inequality increases preferences towards algorithm decision-making. Although our predicted interaction between threat of inequality and personal relevance was not significant, we did find that threat of inequality significantly increased Black participants' preference for algorithms but did not significantly affect White participants' preferences, partially replicating our previous findings. The results of Study 3 also reveal one possible psychological mechanism underlying the effect of threat of inequality on preference for algorithm decision-making. Threat of inequality weakens the perceived authority of doctors and thereby

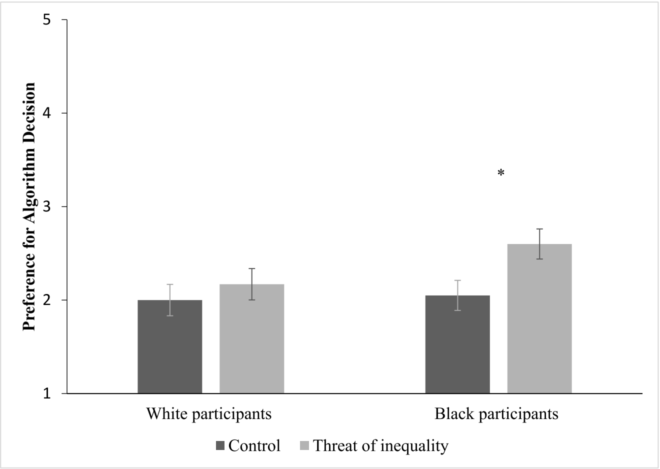

Fig. 3. Preference for Algorithm triage by personal relevance and condition (Study 3). Error bars reflect standard errors.

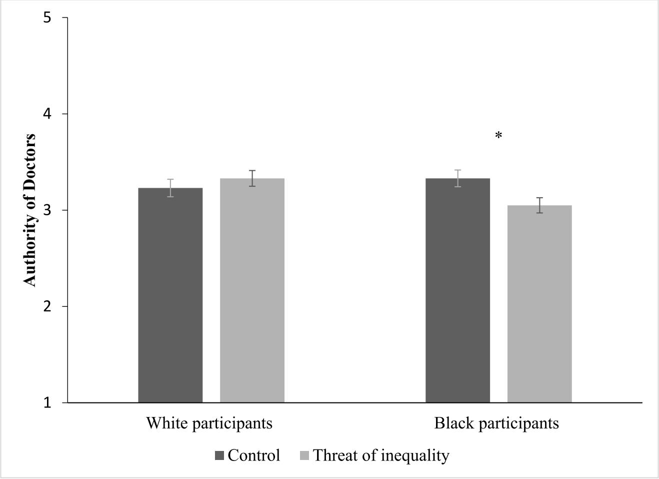

Fig. 4. Belief in doctors' authority by personal relevance and condition (Study 3). Error bars reflect standard errors.

increases people's preference for algorithm decision-making (especially for those disadvantaged by the disparity).

6. Study 4: hospital choice

In Studies 1–3 we measured preference for an algorithm triage on a continuous scale. However, in practice, the choice between hospitals is typically dichotomous. Therefore, in Study 4 we used a similar paradigm to that of Studies 1 and 3 with a binary choice measure, asking participants to choose one of the two hospitals, mirroring real-life decisions. In addition, as in Study 2, we did not mention any possible biases in doctors' decision-making to reduce demand characteristics. We also expanded the groups we describe as having higher mortality rates in our threat of inequality to include Latino and Indigenous people, as well as Black people, to better reflect current data regarding COVID-19 mortality rates (APM Research Lab, 2020). We also explored an additional possible outcome for threat of inequality – support for government assistance to hospitals. Finally, to explore whether SES, rather than race, better captures personal relevance, we also measured SES in this study.

6.1. Method

6.1.1. Participants

An *a priori* power analysis (using G*power 3.1, Faul et al., 2007) revealed that we need a sample size of 800 participants to detect a small effect size. To account for participants who might fail the attention check, we recruited 1005 participants through Prolific. We aimed for an equal number of Black and White participants in our samples (with specified samples). As specified in the pre-registration (https://aspredicted.org/wg8up.pdf), we excluded participants who did not self-report as either Black/Latino/Indigenous or White or those who failed the attention and manipulation comprehension checks. Our final sample was 860 participants (432 male, 419 female, 9 other; 466 White, 190 Hispanic or Latino, 208 Black or African American and 16 Native American or American Indian; Age: $M = 32.08$, $SD = 10.54$). When including all participants in the analysis, the main effect for condition becomes only marginally significant ($p = .073$), the rest of the results remained unchanged. Participants were paid $0.4 as compensation.[5]

6.1.2. SES

As in Study 2, we measured SES with the MacArthur subjective SES scale (Adler et al., 2000).[6]

6.1.3. Threat of inequality manipulation

The control condition was identical to that of Studies 1–3. In the "threat of inequality condition" participants also read the following:

There are large racial and ethnic disparities in who dies from COVID-19. In the United States, for example, the mortality rate from COVID-19 for Black, Latino, and Indigenous people is three times that of White people.

6.1.4. Hospital choice

We asked participants to which hospital they would go, either the hospital where the human doctor makes triage decisions or the hospital where an algorithm makes triage decisions.

6.1.5. Government assistance

We told participants that the government can give medical supplies to one of the two hospitals and asked them which of the two hospitals should receive the medical supplies.

6.1.6. Attention checks

We used the same attention checks as in Study 3.

6.2. Results

Descriptive statistics and correlations are presented in Table 5.

6.3. Choice of hospital

To avoid an interaction term created by multiplying variables that half of their values are equal to 0, we modified the coding of the condition and personal relevance variables for hypothesis testing. A logistic regression with condition (control = −1; threat of inequality = 1), personal relevance (White = −1; Black/Latino/Indigenous = 1), and their interaction predicting choice of hospital revealed a main effect for condition, Wald χ^2 (1, N = 860) = 6.37, $p = .012$, Exp(B) = 1.23, such that participants were more likely to choose algorithm triage in the threat of inequality condition (26.01%) than in the control condition (18.76%), supporting Hypothesis 1. We also found a main effect for

Table 5
Descriptive statistics and correlations for study variables (study 4).

Variable	Mean	SD	1	2	3	4	5
1. Threat of inequality	.45	.50	(- -)				
2. Personal relevance	.48	.50	-.02	(- -)			
3. SES	5.23	1.69	.00	−0.9*	(- -)		
4. Choice of hospital	0.22	0.42	.09*	.09*	-.01	(- -)	
5. Government assistance	0.25	0.44	.08*	.04	.03	.65*	(- -)

Notes.
Threat of inequality: 0 = control condition; 1 = threat of inequality.
Personal relevance: 0 = White participants; 1 = Black/Latino/Indigenous participants.
Choice of hospital: 0 = Hospital with human triage; 1 = Hospital with algorithm triage.
Government assistance: 0 = Hospital with human triage; 1 = Hospital with algorithm triage.
*p < .05.

[5] In Prolific, we could target participants according to race and ethnicity without using a specialized panel, and therefore we could offer participants a more modest compensation.

[6] SES did not interact with condition to predict preference for algorithm decision-making ($p = .162$).

personal relevance, Wald χ^2 (1, N = 860) = 7.47, p = .006, Exp(B) = 1.25, such that Black, Latino, and Indigenous people were more likely to select algorithm triage (26.01%) than White participants (18.39%). However, the personal relevance × condition interaction was not significant, p = .299, which does not support Hypothesis 2. This may reflect the lower statistical power of interaction tests, especially with uneven sample sizes (Simonsohn, 2015).

Although the interaction was not significant, to understand our results better we ran follow-up chi-squared analyses. We found that Black, Latino, and Indigenous people were more likely to choose algorithm triage in the threat of inequality condition (32.79%) than in the control condition (20.96%), χ^2 (1, N = 860) = 6.98, p = .008, ϕ = 0.13. In contrast, for White participants, the difference in choice of algorithm triage between the choice of inequality condition (20.39%) and the control condition (16.67%) was not significant, p = .312, see Fig. 5.

6.4. Government assistance

Logistic regression with condition, personal relevance, and their interaction predicting choice of hospital to receive government assistance revealed a main effect for condition, Wald (1, N = 860) = 4.81, p = .028, Exp(B) = 1.19, such that participants were more likely to choose the hospital with algorithm triage for government assistance in the threat of inequality condition (28.90%) than in the control condition (22.39%). The main effect for personal relevance was not significant, p = .248, and neither was the personal relevance × condition interaction, p = .121.

Although the interaction was not significant, to understand our results better we ran a follow-up chi-squared analysis. We found that Black, Latino, and Indigenous people were more likely to choose algorithm triage for government assistance in the threat of inequality condition (33.51%) than in the control condition (21.38%), χ^2 (1, N = 860) = 7.07, p = .008, ϕ = 0.13. In contrast, for White participants, the difference in choice of algorithm triage between the choice of inequality condition (24.76%) and the control condition (22.92%) was not significant, p = .649.

6.5. Discussion

Relying on a measure of binary choice rather than preferences, the results of Study 4 generalize our earlier findings. Threat of inequality increased the likelihood that participants will choose a hospital where an algorithm makes triage decisions. However, in this forced-choice scenario, we do note that participants still overwhelmingly chose human doctors over algorithms. This suggests that while our manipulation was successful in reducing algorithm aversion, its effect might be small.

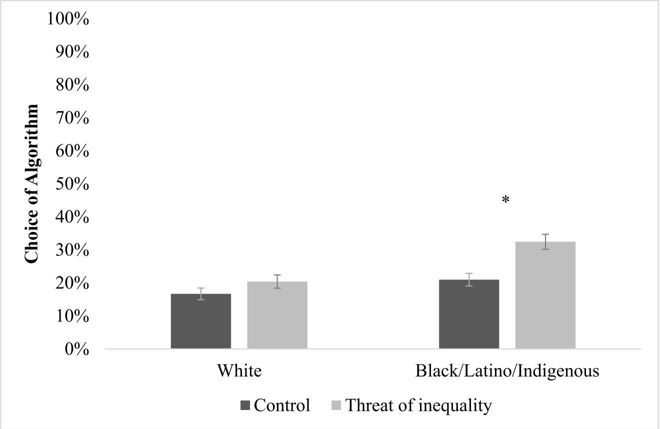

Fig. 5. Choice of hospital with algorithm triage by personal relevance and condition (Study 4). Error bars reflect standard deviations.

In addition, our threat of inequality manipulation led to another high-stake outcome—allocation of government assistance. Although we did not find a significant condition x personal relevance interaction, a simple effect analysis revealed that threat of inequality significantly affected choice of hospital for our Black, Latino and Indigenous participants, but not for our White participants. We also found that threat of inequality affects another applied outcome – it increases people's support for the hospital with algorithm triage to receive government support.

7. Internal meta-analyses

To estimate the overall effect of threat of inequality on preference for algorithm decision-making and to provide a concise summary of our findings (Goh et al., 2016), we conducted internal meta-analyses for Hypotheses 1 and 2. These meta-analyses included the results of Studies 1–3, which used a continuous DV, and were conducted with the 'meta' package in R (Schwarzer, 2007).

The first meta-analysis tested Hypothesis 1 – whether threat of inequality increased preference for algorithm decision-making. This meta-analysis revealed a mean effect size of Cohen's d = 0.42, 95% CI [0.06, 0.78], t = 5.08, p = .037, for a random effect model, and a mean effect size of Cohen's d = 0.45, 95% CI [0.36, 0.54], Z = 9.83, p < .001, for a fixed effect model, supporting Hypothesis 1.The second meta-analysis tested Hypothesis 2 – whether the effect of threat of inequality is moderated by personal relevance. This meta-analysis revealed a mean effect size of Cohen's d = 0.26, 95% CI [-0.03, 0.57], t = 3.84, p = .062, for a random effect model, and a mean effect size of Cohen's d = 0.21, 95% CI [0.15, 0.28], Z = 6.25, p < .001, for a fixed effect model, providing only partial support for Hypothesis 2.

8. General discussion

People are generally averse to algorithms making decisions—at the cost of efficiency and potentially human lives (Bigman & Gray, 2018; Longoni et al., 2019). Making the threat of inequality salient, however, can reduce this aversion. In four studies, we found that the threat of inequality in COVID-19 treatment reduces the aversion people have for algorithms making decisions. This basic pattern replicated across cultures (the United States and Singapore), in various time-points during the pandemic (see Table 1), for a variety of medical decisions such as ventilator allocation and preference for which hospital should receive government assistance, and with (Studies 1 and 3) or without (Studies 2 and 4) the explicit mention of possible bias in human decision-making.

Furthermore, we find some evidence that this aversion reduction is stronger for those to whom the inequality is personally relevant. We note that algorithm aversion for those who personally suffer from human-induced inequality might not be surprising, as such a response could be explained by simple self-interest. Disadvantaged patients might simply prefer the option that will maximize their chances of getting the best medical treatment. However, those who stand to benefit from inequality (e.g., White participants, high-SES participants) also shift their preference toward algorithms, indicating that these changes in preference cannot be solely explained by self-interest.

Mediation analyses help explain the underlying mechanism for this general move in preferences toward algorithms: the reduced aversion from algorithm decision-making is due, at least partially, to a weakening of the perceived authority of doctors. Threat of inequality might cause people to question the authority of doctors—who can be seen as associated with the inequality—and therefore increase people's preference for algorithms, rather than human doctors, as decision-makers. That said, we acknowledge that the empirical support for the role of doctors' authority is limited. Only one study (Study 3) examined this question, and that study explicitly mentioned that doctors might be biased. Future research is needed to test the robustness of this finding.

Our results suggest that people are not only motivated by their self-interest, but also by broader ethical considerations (Jones, 1991; Reynolds, 2006). Our research demonstrates how awareness of the potential race and class inequality in medical decision-making can reduce the aversion from algorithm decision-making and accelerate the integration of algorithms into healthcare settings. By doing so, it can shape the future of human-algorithm interaction in the healthcare system (Hao, 2020). These findings can increase laypeople and patients' acceptance of algorithms and other technologies in the healthcare setting, with the potential to increase efficiency, reduce biases, and even save lives, all of which are valuable implications during and after the COVID-19 pandemic.

To maintain external validity, our studies did not inform participants how the algorithm would make decisions. In practice, when people are informed that a hospital uses algorithms for triage, they do not necessarily receive a description of the decision-making process. Indeed, even the decision-making process of human doctors is often opaque to patients. However, people might have different preferences for different processes of algorithm decision-making (Bonnefon et al., 2016), and general transparency about how AIs make decisions might further increase acceptance of AI decision-making (Kim & Hinds, 2006). Future research is needed to examine this question.

Triage decisions are one domain of high-stake decision-making in which there is a growing public appreciation of outcome inequality, and in which algorithm decision-making is a growing alternative to human decision-making. Other such contexts are banking decisions, such as loans and credit limits (O'neil, 2016; Perez, 2019; Perry, 2019). While our studies focused on medical decision-making, they might extend to other domains, such as banking, in which people might be able to choose whether to apply for a loan from a human banker or an algorithm. Future research is needed to examine the generalizability of our findings to other domains.

We note that while threat of inequality reduced the aversion from algorithm decision-making, preferences for algorithm decision-making were not high. The relatively small effect size is perhaps most salient in Study 4 where participants were forced to choose between either a human or an algorithm. This outcome is consistent with previous work that found a strong aversion against algorithms making moral and medical decisions (e.g., Bigman & Gray, 2018; Longoni et al., 2019). While our short text-based manipulation might not have reversed people's preferences, even small changes in preferences (Studies 1–3) and in choice (Study 4) can yield big differences when aggregated across large populations (Prentice & Miller, 1992). Also, our manipulation is virtually costless while other factors might not be so. That said, future research is needed to examine additional factors that might more dramatically reduce people's algorithm aversion. For example, studies have shown that anthropomorphized algorithms can generally lead to satisfaction in consumer contexts (Rauschnabel & Ahuvia, 2014; Yam et al., 2020), but whether this is the same in the medical context remains an open empirical question. All in all, we are not suggesting that our work can eliminate algorithm aversion, but rather we hope our work will spark additional research to mitigate this aversion.

We acknowledge that although algorithms might be perceived as being more objective than humans (Lee, 2018), several discriminatory algorithms have been documented (O'neil, 2016; Perez, 2019). For example, algorithms can be biased in hiring decisions (Dastin, 2018), parole recommendations (Angwin et al., 2016), and identifying people in need of special medical assistance (Obermeyer et al., 2019). Although some algorithms might be biased, algorithm bias is often easier to be corrected than human bias (Mullainathan, 2019). We also acknowledge that structural inequalities, and not necessarily individual bias, could be a major source of the higher COVID-19 mortality rates of some groups (Abuelgasim et al., 2020; Bibbins-Domingo, 2020; Braveman et al., 2011; Yancy, 2020). We note that it is possible (and plausible) that learning about algorithm bias might increase people's aversion to algorithm decision-making. However, at least currently, the role of algorithm decision-making in medical decision-making is still limited, and people are less likely to perceive them as biased (Bigman et al., 2020; Lee, 2018).

Replacing human decision-making with algorithm decision-making will not eliminate health disparity. Access to quality healthcare, habitat density, the ability to work remotely, and SES all contribute to the higher mortality rate in some communities (Abuelgasim et al., 2020; Bibbins-Domingo, 2020; Braveman et al., 2011; Yancy, 2020). Although we demonstrated that the threat of inequality can increase preference for algorithm decision-making, this does not necessarily translate to reduced inequality in medical outcomes. Future work should consider all these factors in tandem with human vs. algorithm-based medical decision-making. We do hope, however, that our work can contribute to this ongoing discussion about healthcare disparities and offer one way to potentially mitigate them.

8.1. *Limitations and future directions*

We note that there is some variability in the effect sizes for our threat of inequality manipulation. Specifically, the effect size for threat of inequality in Study 1 ($\eta_p^2 = .07$) is much larger than in Study 3 ($\eta_p^2 = .02$), although they both used the same manipulation with the same population. One possibility is that the demonstrations following the killing of George Floyd in the U.S. on May 25, 2020 mitigated the effectiveness of our manipulation, as the baseline of threat of inequality might have changed. Another possibility is that as time passed and countries equipped themselves better, the fear of overcrowded hospitals and the lack of ventilators was reduced. Future research is needed to explore these possibilities. The differences in effect sizes in our other studies are not surprising, as they included samples from different populations, with different manipulations and different dependent variables.

One limitation of our studies is that our main dependent variables are self-reported preferences and decisions, which although being a reliable measure of general attitudes (Krosnick et al., 2005) is not as compelling as a field study measuring actual behaviors. However, ethical considerations preclude the possibility of such a field study, as it would involve asking patients to select either a human doctor or an algorithm as they are being admitted into a hospital with acute COVID-19 symptoms when rapid treatment is critical. Nevertheless, future research should examine this question in a field study, perhaps among patients with less acute illnesses.

In our studies, we simplified the medical decision-making process and the role of AI in this process. In actual healthcare settings, decision-making is complex and often involves several parties and stakeholders. These can be the various nurses, doctors, the patient, and the patient's family. While AI input might be one factor that determines treatment, at least today, it is not the only one. To capture this complexity, future studies could ask participants to choose between a hospital staffed by only human doctors, nurses, and healthcare professionals vs. a hospital staffed by human doctors, nurses, healthcare professionals, *and* AI.

In life-and-death medical decisions, the need for accountability and justification is high, as they involve legal liability. This might also be an obstacle in the dissemination of deep learning algorithms in which the decision-making process is not always fully explainable (Abdul et al., 2018; Alaieri & Vellino, 2016). These limitations should be taken into account when considering the applications of this research.

Another limitation of our studies is the operationalization of medical decision-making as triage decisions when resources are scarce. These decisions might be specific to pandemics, for which the healthcare system is not properly equipped and prepared. Future research should examine people's preferences for a wider range of decisions by algorithms. For example, as of this writing, vaccine allocation has been a heatedly debated topic and future research can explore whether our findings can be generalized to this specific outcome.

8.2. Concluding remarks

The use of algorithms for decision-making holds promise to help society in many ways (Jackson et al., 2020), including increasing efficiency and fairness in healthcare. Here we show how awareness of health disparities can increase people's acceptance of algorithms and accelerate the integration of algorithms into the workforce, with potential lasting impacts for racial and economic equality.

Credit author statement

Yochanan E. Bigman: Conceptualization, Methodology, Investigation, Analysis, Writing - Original Draft, Review and Editing. Kai Chi Yam: Conceptualization, Methodology, Investigation, Analysis, Writing - Review and Editing, Funding Acquisition. Déborah Marciano: Conceptualization, Methodology, Writing - Review and Editing. Scott J. Reynolds: Methodology, Writing - Review and Editing. Kurt Gray: Conceptualization, Methodology, Analysis, Writing - Review and Editing, Funding Acquisition.

Decaration of competing interest

We have no interest of conflict to disclose.

Acknowledgement

This research was funded by the Singapore Ministry of Education Tier 1 Grant (R-317-000-149-115), and by the Charles Koch Foundation via the Center for the Science of Moral Understanding.

References

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. *Conference on Human Factors in Computing Systems - Proceedings, 2018-April*, 1–18. https://doi.org/10.1145/3173574.3174156

Abuelgasim, E., Saw, L. J., Shirke, M., Zeinah, M., & Harky, A. (2020). COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities. *Current Problems in Cardiology, 45*(8). https://doi.org/10.1016/j.cpcardiol.2020.100621, 100621.

Adler, N. E., Epel, E. S., Castellazzo, G., & Ickovics, J. R. (2000). Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy white women. *Health Psychology, 19*(6), 586–592. https://doi.org/10.1037/0278-6133.19.6.586

Alaieri, F., & Vellino, A. (2016). Ethical decision making in robots: Autonomy, trust and responsibility autonomy trust and responsibility. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9979 LNAI*, 159–168. https://doi.org/10.1007/978-3-319-47437-3_16

Alsan, M., Garrick, O., & Graziani, G. (2019). Does diversity matter for health? Experimental evidence from oakland. *The American Economic Review, 109*(12), 4071–4111. https://doi.org/10.1257/aer.20181446

Angwin, J., Larson, J., Surya, M., & Lauren, K. (2016). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

APM Research Lab. (2020). The color of coronavirus: COVID-19 deaths by race and ethnicity in the U.S. APM research Lab. https://www.apmresearchlab.org/covid/deaths-by-race.

Ardon, O., & Schmidt, R. L. (2020). Clinical laboratory employees' attitudes toward artificial intelligence. *Laboratory Medicine, 51*(6), 649–654. https://doi.org/10.1093/labmed/lmaa023

Arias, E., Tejada-vera, B., & Ahmad, F. (2021). Provisional life expectancy estimates for January through June, 2020. *CDC Vital Statistics Rapid Release, 10*, 1–8. https://www.cdc.gov/nchs/data/vsrr/VSRR10-508.pdf?utm_source=STAT+Newsletters&utm_campaign=63a6765dc6-MR_COPY_14&utm_medium=email&utm_term=0_8cab1d7961-63a6765dc6-149545845.

Bauchner, H., & Easley, T. J. (2020). Health care heroes of the COVID-19 pandemic. *JAMA, 323(20), 2021*. https://doi.org/10.1001/jama.2020.6197

Bhala, N., Curry, G., Martineau, A. R., Agyemang, C., & Bhopal, R. (2020). Sharpening the global focus on ethnicity and race in the time of COVID-19. *The Lancet, 395*, 1673–1676. https://doi.org/10.1016/S0140-6736(20)31102-8, 10238.

Bibbins-Domingo, K. (2020). This time must Be different: Disparities during the COVID-19 pandemic. Annals of internal medicine, M20–2247. https://doi.org/10.7326/M20-2247.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Bigman, Y. E., Wilson, D., Arnestad, M. N., Waytz, A., & Gray, K. (2020). *Algorithmic DiscriminationCauses less moral outrage than human discrimination*.

Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial intelligence and the future of primary care: Exploratory qualitative study of UK general practitioners' views. *Journal of Medical Internet Research, 21*(3), e12802. https://doi.org/10.2196/12802

Board of Governors of the Federal Reserve System. (2021). Share of total net worth held by the top 1% (99th to 100th wealth percentiles). https://fred.stlouisfed.org/graph/?graph_id=710397&rn=363.

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Brase, G. L., & Richmond, J. (2004). The white–coat effect: Physician attire and perceived authority, friendliness, and attractiveness. *Journal of Applied Social Psychology, 34*(12), 2469–2481. https://doi.org/10.1111/j.1559-1816.2004.tb01987.x

Braveman, P., Egerter, S., & Williams, D. R. (2011). The social determinants of health: Coming of age. *Annual Review of Public Health, 32*, 381–398. https://doi.org/10.1146/annurev-publhealth-031210-101218

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research, 56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Crouch, H. (2019). London hospital trials digital triage service in urgent care departments. DigitalHealth. https://www.digitalhealth.net/2019/06/london-hospital-trials-digital-triage/.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

Devlin, H., & Barr, C. (2020). Poorest areas of england and wales hit hardest by covid-19 – ONS. The guardian. https://www.theguardian.com/world/2020/jun/12/poorest-areas-of-england-and-wales-hit-hardest-by-covid-19-ons.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Doraiswamy, P. M., Blease, C., & Bodner, K. (2020). Artificial intelligence and the future of psychiatry: Insights from a global physician survey. *Artificial Intelligence in Medicine, 102*. https://doi.org/10.1016/j.artmed.2019.101753, 101753.

Dubois, D., Rucker, D. D., & Galinsky, A. D. (2015). Social class, power, and selfishness: When and why upper and lower class individuals behave unethically. *Journal of Personality and Social Psychology, 108*(3), 436–449. https://doi.org/10.1037/pspi0000008

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12*(1), 1–22. https://doi.org/10.1037/1082-989X.12.1.1

Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race–ethnicity, and sex. *Proceedings of the National Academy of Sciences of the United States of America, 116*(34), 16793–16798. https://doi.org/10.1073/pnas.1821204116

Eligon, J., Burch, A. D. S., Searcey, D., & Oppel, R. A. J. (2020). Black Americans face alarming rates of coronavirus infection in some States. The New York times. https://www.nytimes.com/2020/04/07/us/coronavirus-race.html.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition. Mcgraw-hill book company*.

Gauher, S., & Uz, F. B. (2016). Cleveland clinic to identify at-risk patients in ICU using cortana intelligence. Machine learning blog. https://docs.microsoft.com/en-us/archive/blogs/machinelearning/cleveland-clinic-to-identify-at-risk-patients-in-icu-using-cortana-intelligence-suite.

Geddi, J., Brock, J., & Koustav, S. (2020). Singapore's migrant workers fear financial ruin after virus ordeal. *Rueters*. https://www.reuters.com/article/us-health-coronavirus-singapore-migrants/singapores-migrant-workers-fear-financial-ruin-after-virus-ordeal-idUSKBN23G1PG.

Gerretsen, I. (2020). Robots are joining the fight against coronavirus in India. CNN Business. https://www.cnn.com/2020/11/11/tech/robots-india-covid-spc-intl/index.html.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). *Mini Meta-Analysis of Your Own Studies : Some Arguments on Why and a Primer on How, 10*, 535–549.

Goranson, A., Sheeran, P., Katz, J., & Gray, K. (2020). Doctors are seen as Godlike: Moral typecasting in medicine. *Social Science & Medicine, 113008*. https://doi.org/10.1016/j.socscimed.2020.113008

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366–385. https://doi.org/10.1037/a0021847

Hao, K. (2020). Doctors are using AI to triage covid-19 patients. The tools may be here to stay. *MIT Technology Review*, 1–12. https://www.technologyreview.com/2020/04/23/1000410/ai-triage-covid-19-patients-health-care/.

Hoffman, K. M., Trawalter, S., Axt, J. R., & Oliver, M. N. (2016). Racial bias in pain assessment and treatment recommendations, and false beliefs about biological differences between blacks and whites. *Proceedings of the National Academy of Sciences of the United States of America, 113*(16), 4296–4301. https://doi.org/10.1073/pnas.1516047113

Houser, K. A. (2019). Can AI solve the diversity problem in the tech industry? Mitigating noise and bias in employment decision-making. *Stanford Technology Law Review, 22*, 1–42.

Jackson, J. C., Castelo, N., & Gray, K. (2020). Could a rising robot workforce make humans less prejudiced? *American Psychologist, November*. https://doi.org/10.1037/amp0000582

Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review, 16*(2). https://doi.org/10.2307/258867, 366.

Kim, T., & Hinds, P. (2006). Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*, 80–85. https://doi.org/10.1109/ROMAN.2006.314398

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). *Measurement of attitudes. Handbook of attitudes and attitude change, 21–76*.

Laï, M.-C., Brian, M., & Mamzer, M.-F. (2020). Perceptions of artificial intelligence in healthcare: Findings from a qualitative survey study among actors in France. *Journal of Translational Medicine, 18*(1). https://doi.org/10.1186/s12967-019-02204-y, 14.

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science, 65*(7), 2966–2981. https://doi.org/10.1287/mnsc.2018.3093

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data and Society, 5*(1), 1–16. https://doi.org/10.1177/2053951718756684

Liang, W., Yao, J., Chen, A., Lv, Q., Zanin, M., Liu, J., Wong, S. S., Li, Y., Lu, J., Liang, H., Chen, G., Guo, H., Guo, J., Zhou, R., Ou, L., Zhou, N., Chen, H., Yang, F., Han, X., et al.He, J. (2020). Early triage of critically ill COVID-19 patients using deep learning. *Nature Communications, 11*(1), 1–7. https://doi.org/10.1038/s41467-020-17280-8

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime. Com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. https://doi.org/10.3758/s13428-016-0727-z, 433–442.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research, 46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Mangum, M. (2008). Testing competing explanations of black opinions on affirmative action. *Policy Studies Journal, 36*(3), 347–366. https://doi.org/10.1111/j.1541-0072.2008.00267.x

McCall, L., Burk, D., Laperrière, M., & Richeson, J. A. (2017). Exposure to rising inequality shapes Americans' opportunity beliefs and policy support. *Proceedings of the National Academy of Sciences of the United States of America, 114*(36), 9593–9598. https://doi.org/10.1073/pnas.1706253114

Meah, N. (2020). Robot to deliver meals, medication to Covid-19 patients at Alexandra Hospital to reduce exposure of healthcare workers Read more at. Today https://www.todayonline.com/singapore/robot-deliver-meals-medication-covid-19-patients-alexandra-hospital-reduce-exposure https://www.todayonline.com/singapore/robot-deliver-meals-medication-covid-19-patients-alexandra-hospital-reduce-exposure.

Morrison, S. R., Wallenstein, S., Natale, D. K., Senzel, R. S., & Haung, L.-L. (2000). " we don't carry that " — failure of pharmacies in predominantly nonwhite neighborhoods to stock opioid analgesics. *New England Journal of Medicine, 342*(14), 1023–1026.

Mullainathan, S. (2019, December 6). Biased Algorithms Are Easier to Fix Than Biased People. The New York Times. https://www.nytimes.com/2019/12/06/business/algorithm-bias-fix.html.

Munoz, C., Smith, M., & Patil, D. (2016). Big data: A report on algorithmic systems, opportunity, and civil rights big Data : A report on algorithmic systems, opportunity, and civil rights. *Executive Office of the President of USA, May*. https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

Murphey, S. L., Xu, J., Kochanek, K. D., Arias, E., & Tekada-Vera, B. (2021). Deaths: Final data for 2018. National vital statistics reports. http://www.mendeley.com/research/deaths-final-data-for-2002-national-vital-statistics-reports/, 69, 13.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

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

Oxford University Press. (2020). Definition of bias. In *Lexico.com*. https://www.lexico.com/en/definition/bias.

Palma. (2020). Surge in Covid cases shows up Singapore's blind spots over migrant workers. *Financial Times*. https://www.ft.com/content/0fdb770a-a57a-11ea-92e2-cbd9b7e28ee6.

Parrock, J. (2020). Coronavirus: Belgium hosptial employs robot to protect against COVID-19. Euronews. https://www.euronews.com/2020/06/02/coronavirus-belgium-hospital-employs-robot-to-protect-against-covid-19.

Perez, C. C. (2019). *Invisible women: Exposing data bias in a world designed for men*. Random House.

Perry, V. G. (2019). A loan at last? Race and racism in mortgage lending. In G. D. Johnson, K. D. Thomas, A. K. Harrison, & S. A. Grier (Eds.), *Race in the marketplace: Crossing critical boundaries* (pp. 173–192). Springer International Publishing. https://doi.org/10.1007/978-3-030-11711-5_11.

Piff, P. K., Stancato, D. M., Coteb, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences of the United States of America, 109*(11), 4086–4091. https://doi.org/10.1073/pnas.1118373109

Polesie, S., Gillstedt, M., Kittler, H., Lallas, A., Tschandl, P., Zalaudek, I., & Paoli, J. (2020). Attitudes towards artificial intelligence within dermatology: An international online survey. *British Journal of Dermatology, 183*(1), 159–161. https://doi.org/10.1111/bjd.18875

Prentice, D. A., & Miller, D. T (1992). When small effects are impressive. *Psychological Bulletin, 112*, 160–164. https://doi.org/10.1037/0033-2909.112.1.160.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Blenknap.

Rauschnabel, P. A., & Ahuvia, A. C. (2014). You're so lovable: Anthropomorphism and brand love. *Journal of Brand Management, 21*(5), 372–395. https://doi.org/10.1057/bm.2014.14

Rest, J. R. (1986). *Moral development: Advances in research and theory*. Praeger Publishers.

Reynolds, S. J. (2006). Moral awareness and ethical predispositions: Investigating the role of individual differences in the recognition of moral issues. *Journal of Applied Psychology, 91*(1), 233–243. https://doi.org/10.1037/0021-9010.91.1.233

Reynolds, S. J. (2008). Moral attentiveness: Who pays attention to the moral aspects of life? *Journal of Applied Psychology, 93*(5), 1027–1041. https://doi.org/10.1037/0021-9010.93.5.1027

Reynolds, S. J., & Miller, J. A. (2015). The recognition of moral issues: Moral awareness, moral sensitivity and moral attentiveness. *Current Opinion in Psychology, 6*, 114–117. https://doi.org/10.1016/j.copsyc.2015.07.007

Sands, M. L. (2017). Exposure to inequality affects support for redistribution. *Proceedings of the National Academy of Sciences of the United States of America, 114*(4), 663–668. https://doi.org/10.1073/pnas.1615010113

Schwarzer, G. (2007). meta: An R package for meta-analysis. *R News, 7*(3), 40–45.

Shea, G. P., Laudanski, K., & Solomon, C. A. (2020). Triage in a pandemic: Can AI help ration access to care?. Knowledge@Wharton. https://knowledge.wharton.upenn.edu/article/triage-in-a-pandemic-can-ai-help-ration-access-to-care/.

Simonsohn, U. (2015). [17] No-way interactions. The winnower. https://doi.org/10.15200/winn.142559.90552.

Tayarani-N, M.-H. (2020). Applications of artificial intelligence in battling against covid-19: A literature Review. *Chaos, Solitons & Fractals, 110338*. https://doi.org/10.1016/j.chaos.2020.110338

Thom, D. H., Ribisl, K. M., Stewart, A. L., & Luke, D. A. (1999). Further validation and reliability testing of the trust in physician scale. *Medical Care, 37*(5), 510–517. https://doi.org/10.1097/00005650-199905000-00010

Vanian, J. (2020). How chatbots are helping in the fight against COVID-19. Fortune. https://fortune.com/2020/07/15/covid-coronavirus-artificial-intelligence-triage/.

Yam, K. C., Bigman, Y. E., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2020). Robots at work: People prefer-and forgive-service robots with perceived feelings. *Journal of Applied Psychology*. https://doi.org/10.1037/apl0000834

Yancy, C. W. (2020). COVID-19 and african Americans. *JAMA - Journal of the American Medical Association, 323*(19), 1891–1892. https://doi.org/10.1001/jama.2020.6548

Young, A. D., & Monroe, A. E. (2019). Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. *Journal of Experimental Social Psychology, 85*. https://doi.org/10.1016/j.jesp.2019.103870. August 2018.