Slower Moral Tradeoffs by AI Enhance AI Appreciation

Adelle X. Yang

(adelle.yang@gmail.com)

Assistant Professor, National University of Singapore, Singapore

Sijin Chen

(chensijin@u.nus.edu)

PhD Candidate, National University of Singapore, Singapore

and

Yu Gu (guy.18@sem.tsinghua.edu.cn)

Assistant Professor, Nankai University, China

Manuscript in preparation, Oct 2024

Slower Moral Tradeoffs by AI Enhance AI Appreciation

Abstract (187/200)

While speed is often considered a key advantage of Artificial Intelligence (AI), our research challenges this assumption in morally complex contexts. Across 13 pre-registered experiments (N = 8,473), we find that participants rate an AI more favorably when it takes longer to make moral-tradeoff decisions, even when the outcomes remain unchanged. This preference for a slower AI was observed in both classic moral dilemmas and resource-allocation decisions involving individuals in need. Both the "moral" and "tradeoff" elements of the focal decision were necessary for the effect. The effect did not emerge when the AI was slower on a more important non-moral decision. We find that the effect is primarily driven by overgeneralized moral intuitions regarding the AI's decision-making procedure, rather than beliefs about the decision outcomes. While anthropomorphism and mind perception may also partially contribute to the effect, they could not fully account for the effect. Furthermore, analyses of open-ended text responses with a large language model (GPT-4o) corroborate measured process evidence. This research offers new insights into AI resistance, especially in moral domains where AI faces the strongest resistance, and suggests new directions for policy interventions.

Keywords: Artificial Intelligence (AI), moral decisions, moral intuitions, decision speed

Speed is one of the most lauded strengths of Artificial Intelligence (AI). Developers continuously strive to build faster algorithms and AI systems (Mankowitz et al., 2023; OpenAI, 2024), which are valued by consumers (Goodman & Spence, 1978; Xie et al., 2023). Although AI still faces considerable societal resistance (e.g., Narayan et al., 2023; Feiner, 2023), much research has attributed this resistance to people's concerns about AI causing negative consequences such as threatening human autonomy and perpetuating errors and systematic biases (see reviews in Morewedge et al., 2024; Ukanwa 2024; Valenzuela et al., 2024). In contrast, AI's fast speed has been viewed as an unambiguous advantage (Dabrowski & Munson, 2011; Efendić et al., 2020).

However, do people always prefer AI's fast speed? We question the generality of this preference. In particular, we propose that faster decision speed by AI may sometimes decrease evaluation of the AI—even when decision outcomes remain unchanged—such as when the decision involves a moral tradeoff. If this is the case, AI's ostensible speed advantage could be an overlooked cause for the widely documented negative sentiments toward AI, particularly pronounced in morally relevant domains (e.g., Bigman & Gray, 2018; Dietvorst & Bartels, 2022; Jauernig et al., 2022).

We tested this proposal in 13 pre-registered experiments with 8,473 participants from North America and Asia, six of which (N = 2,054) are presented in this paper. The proposed effect was tested with various moral-tradeoff decisions, including the trolley problems (Foot, 1978; Thomson, 1985), tragic tradeoffs (Tetlock et al., 2000), the autonomous vehicle dilemma (Bonnefon et al., 2016), and more generic resource-allocation decisions among individuals in need (e.g., involving healthcare or educational resources). Across diverse contexts, we found consistent evidence that slower moral tradeoffs by AI increase AI appreciation. Moreover, these findings are primarily explained by moral intuitions (i.e., strong and instinctive feelings of right and wrong; Cushman, 2013; Kahneman & Sunstein, 2007), which are sensitive to procedural cues, such as decision speed. Critically, rapid moral tradeoffs by AI violate expectations about how morally complex decisions *ought to* be made, thereby eliciting negative moral intuitions regarding the AI's moral desirability. Even though people understand that AI does not possess moral intentions, these moral intuitions dampen people's evaluation of the AI independently from consequentialist considerations. We also found anthropomorphism and related mind perception could partially contribute to the effect but could not fully account for it. Our findings shed new light on the root causes of AI aversion and suggest novel interventions for mitigating it.

AI Decision-Making: Is Faster Always "Better"?

How does decision speed impact people's evaluation of AI and advanced algorithms? At first glance, a consensus appears to emerge across domains: the faster, the better. Tech leaders such as Google and Bing have found that users exhibit a strong preference for faster responses from algorithmic search tools—even minimal delays in search results can significantly reduce the number of searches per user (Shurman & Brutlag, 2009). Behavioral research has also reported a general preference for faster AI (Efendić et al., 2020; Xie et al., 2023; see also Dabrowski & Munson, 2011). For example, Efendić et al. (2020) showed that consumers preferred faster algorithms and avoided slower ones in computational and forecasting tasks. Because an AI's decision speed is a direct indicator of its computational prowess, it is reasonable that people often judge a slow AI negatively as if less capable and less useful. This preference for speed contrasts with the evaluation of human decision-making, where slower human decisions on difficult choices are rather appreciated for signaling a more careful and thorough underlying decision process (Kupor et al., 2014; also see Landy et al. 2024).

Admittedly, fast decisions are technically desirable; they improve efficiency and can sometimes save lives (e.g., earthquake alerts; Hutchinson, 2023). But how people evaluate

AI's desirability is not always determined solely by its technical merits. While the preference for faster AI speed in computational and forecast tasks is uncontroversial, it is important to examine whether this preference generalizes to other domains.

Today, as AI is increasingly used in decision systems that involve complex moral tradeoffs, including in self-driving cars (Nvidia, 2017) and policy systems such as those distributing social welfare and healthcare (Stone et al., 2016), consumers are inevitably faced with purchase decisions and policy systems involving AI systems with direct and indirect moral implications. Their preference and evaluation of the AI will ultimately affect the success of marketing activities and policy proposals How does AI's speed affect people's evaluation of the AI in these morally complex decisions? If AI's faster speed hurts, rather than helps, then keep emphasizing AI's speed advantage would backfire and potentially exacerbate AI aversion rather than mitigate it.

Moral Intuitions and Speed Cues

We propose that how people evaluate an AI decision-maker's desirability may be affected by moral intuitions, which are sensitive to situational factors surrounding the decision-making process (Cushman, 2013), such as its speed. In dual-process models of moral judgments (Cushman et al., 2010; Greene et al., 2004; Greene & Haidt, 2002), a moral actor's desirability is jointly determined by moral reasoning and moral intuitions. Like other forms of consequentialist thinking, moral reasoning is analytical, calculative, and outcome-oriented (Cushman, 2013; Greene & Haidt, 2002; Kohlberg, 1969). In contrast, moral intuitions are deeply felt convictions of right and wrong, which are instinctive, associative, and action-focused (Cushman, 2013; Sinnott-Armstrong et al., 2010). When people have a moral intuition, they often feel as if it is self-evident, universal, and non-negotiable (Kahneman & Sunstein, 2007).

Moral intuitions are often rooted in experience and norms (Cushman 2013). As such, negative moral intuitions can arise simply when a moral action violates expectations about how the action is usually performed, where experience and norms serve as a reference state (Kahneman & Sunstein, 2007). Many also consider moral intuitions as mostly deontological in nature, namely, they are rule-based; they judge an action by itself and largely disregard its consequences (Bartels et al., 2015; Bartels & Medin, 2007). For example, Haidt (2001) had documented that people often object to an action (e.g., consensual sex between adult siblings) based on moral intuitions (e.g., "It just feels morally wrong"), which they not only struggle to justify with reasons, but tend to maintain their objections even after their reasons had been refuted (Haidt, 2001; Haidt et al., 2000).

We propose that AI's decision speed may elicit different moral intuitions, contributing to positive or negative evaluation of the AI as a moral decision-maker. This proposal is built on two premises. First, there could be existing speed-associated moral intuitions that people have acquired from observing human decision-making in morally complex situations. Second, these moral intuitions could be overgeneralized to the AI as a decision-maker, even though people understand that AI does not have morality.

Indeed, the first premise has received some tentative support in prior research on how decision speed affects judgments of human decision-makers (Critcher et al., 2013; Kupor et al., 2014; Van de Calseyde et al., 2014). For example, Tetlock et al., (2000) reported that their participants judged a person who slowly and deliberatively decided on a tragic tradeoff (e.g., choosing which of two boys would receive an organ transplant) more favorably than one who quickly decided, describing the slower decision as if reflecting "*greater respect shown for the solemnity of the decision*" (Tetlock et al., 2000, p. 860). Separately, two unpublished studies by Landy et al. found that individuals who made difficult moral decisions slowly and thoughtfully were judged as having better moral character than those who did so quickly and unthinkingly (Studies 2 & 3 in Landy et al., 2024).

However, because the length and depth of contemplation were always manipulated simultaneously in these studies, it remains unclear whether and how decision speed alone would affect evaluation of the human decision-maker. To conclusively examine this relationship, we conducted a pilot study (*N* = 303; see S1 in Supplemental Materials). In this study, we manipulated decision speed in three between-participants conditions (Moral-Faster vs. Same-Time vs. Moral-Slower), where the person was faster, equal, or slower with a moral tradeoff than a structurally similar non-moral decision (see Study 1 for the same paradigm). We found that the person who took longer to resolve the moral tradeoff was rated the most positively ($M_{\text{Moral-Faster}} = .54$ vs. $M_{\text{Same-Time}} = .68$ vs. $M_{\text{Moral-Slower}} = 1.73$; $F(2, 300) = 8.25$, $p < .001$, $\eta_p^2 = .05$), even though the person's decision was expected to be similar across conditions ($\chi^2(2, N = 303) = 1.06$, $p = .59$). These results indicate that the relative speed of a moral tradeoff can affect the desirability judgment of human decision-makers.

Then, will speed cues similarly affect the desirability judgment of an AI? Moral intuitions have been theorized to be remarkably prone to overgeneralizations (Cushman, 2013; Kahneman & Sunstein, 2007; Sinnott-Armstrong et al., 2010). Moreover, although moral intuitions are rooted in human interactions (Gray et al., 2012), they are not necessarily exclusive to human actions. Rather, once action cues are associated with moral judgments, they become detached from specific contexts and outcomes (Crockett, 2013; Cushman, 2013). Therefore, we expect that similar moral intuitions may be triggered by AI decision-makers' speed cues—even though it is well-understood that AI does not possess moral intentions, good or bad, irrespective of its length of decision-making.

More specifically, we propose that a rapid moral tradeoff by AI is likely to elicit an overgeneralized norm violation, making the decision procedure feel somewhat disrespectful and objectionable. In contrast, a slower moral tradeoff by AI is perhaps less likely to elicit the same degree of negative moral intuitions. As a result, we expect that slower moral tradeoffs by AI will evoke relatively more positive moral intuitions regarding the decision procedure. Our main hypothesis on the speed effect (H1) is primarily based on this hypothesized relationship between decision speed and moral intuitions (H2).

In summary, we sought to examine how procedural cues affect AI aversion through evoking different moral intuitions, even when expected decision outcomes are unchanged. This novel perspective contrasts with prior research, which has primarily linked AI aversion to various consequentialist concerns regarding the inferiority of AI-produced decision outcomes (e.g., Castelo et al., 2019; Dietvorst et al., 2015; Dietvorst & Bharti, 2020), including in morally relevant domains (Bigman & Gray, 2018; Dietvorst & Bartels, 2022; Young & Monroe, 2019). If moral intuitions drive AI aversion independently from potential consequentialist beliefs, this finding has important implications for current assumptions regarding AI aversion and corresponding interventions.

Two Necessary Conditions

We propose that both “moral” and “tradeoff” in the focal decision are necessary for the proposed effect. If either is removed, the effect should diminish. First, not all moral decisions involve a tradeoff. Moral duties, or moral obligations, are decisions where common moral principles clearly stipulate which outcome is morally acceptable versus not. For example, stealing or cheating are universally condemnable, whereas refraining from them is morally righteous. In these cases, Critcher et al. (2013) have shown that a slow decision to execute moral duties dampens evaluation of the decision-maker because it is seen as revealing the decision-maker’s undesirable moral character. Notably, Critcher et al. (2013) also acknowledged that they expect different results if moral duties were replaced with difficult moral dilemmas (Critcher et al., 2013, p. 7). This conjecture is consistent with theorizing in Tetlock et al. (2000) and supported by our pilot study (S1, SOM): a person who resolves a difficult moral dilemma slower is deemed more positively. Likewise, we do not expect the proposed effect to remain when AI involves a moral duty instead of a moral tradeoff. The type (moral tradeoff vs. moral duty) of the focal decision should moderate how decision speed affects the evaluation of AI.

Second, based on our theorizing on the role of moral intuitions, the proposed effect should no longer appear when the "moral" element is removed. A moral tradeoff involves conflicting moral values, and resolving it inevitably leads to some morally undesirable effects (Bartels et al., 2015; Bartels & Medin, 2007). But not all difficult or important tradeoffs have moral relevance. For example, an investment decision between two profitable funds might involve complex tradeoffs and substantive financial consequences, while lacking any moral implications. Then, to what extent does the speed on such decisions evoke moral intuitions, if at all? We do not expect the same moral intuitions to be associated with non-moral decisions, at least not to the same degree.

However, our proposed effect may be found in such morally irrelevant domains if an alternative "computational hierarchy" account explains the proposed effect: perhaps people view slower decision speed on a focal decision as a signal that the decision received more computational resources, and perhaps people simply prefer AIs that allocate more computational resources to more important decisions. If this is the case, then the proposed "slower is better" effect should be observed as long as the more important decision is slower, even when morally irrelevant.

Finally, theories of mind may also contribute to the proposed effect, besides moral intuitions. People often ascribe minds to non-humans including intelligent robots (Gray et al., 2007; Waytz et al., 2014), a tendency known as anthropomorphism (Epley et al., 2007). From this perspective, decision speed might serve as a cue of resemblance, influencing how human-like an AI is perceived. Some studies have shown that when nonhuman entities exhibit similar movement speed as humans, they are perceived to have more human-like mental states (C. Morewedge et al., 2007). Moreover, assuming that people generally prefer a more human-like AI (de Visser et al., 2016; Waytz et al., 2014), different degrees of anthropomorphism may also contribute to the proposed "slower is better" effect.

These two primary mechanisms are not mutually exclusive, as they operate via largely orthogonal psychological pathways. Moral intuitions can be evoked by familiar procedural cues (e.g., speed) irrespective of the actor's entity, without requiring theories of mind as an antecedent. Nor do theories of mind predict the direction or degree of moral intuitions. Meanwhile, both could contribute to the effect. Therefore, we will separately examine both sets of mechanisms, empirically assessing how much each set contributes to the proposed effect. Additionally, even when situational cues prompt similar levels of anthropomorphism across conditions, we would still expect moral intuitions to hold and the "slower is better" effect to persist. We will also test this proposal to disentangle the primary mechanisms.

Experiments

We pre-registered all studies and reported all manipulations, measures, and analyses. We report all primary results in the main text. We summarize secondary analyses and results in the main text and report their full details in the Supplementary Online Materials (SOM). We share all data, unabridged survey materials, and the SOM on Open Science Framework (anonymized link for review: https://osf.io/ts8d7/?view_only=8db0b5aa5dec4e8d9d00521b84f29f3e).

Sampling and Screening

We conducted 13 pre-registered experiments with 8,473 participants from North America and Asia. We present six of them (N = 3,058) in the main text, and report the rest in the SOM. All studies had target sample sizes of at least 100 participants per between-participants condition. Sample sizes were increased in the subject-pool study (Study 1) and were quadrupled in most studies testing for potential moderations (Studies 4 & S4-S7). All studies with online samples only excluded incomplete responses and duplicate IP addresses, as pre-registered, with similar attrition rates across conditions (Table S1, SOM). Analyses were conducted only after completion of screening procedures.

Overview of Studies

We used a speed-variation paradigm and a choice paradigm. Most studies followed the speed-variation paradigm, in which each participant was presented with an AI decision-maker that was faster, of the same speed, or slower at resolving a moral tradeoff than a non-moral decision of a similar structure. Study 1 used a pair of trolley problems to test which of the three AI induced the most positive impression (-5 = very negative, 0 = neutral, 5 = very positive). Study 2 replicated the "slower is better" effect with resource-allocation decisions among needy individuals. Then, in Study 3, we used the choice paradigm to show that participants preferred a generally slower AI to a generally faster AI in a binary choice only when the slower AI was extra slower on moral tradeoffs, not when it was extra slower on non-moral tradeoffs or moral duties. Study 4 showed that the "slower is better" effect did not emerge when the AI was slower on a more important financial decision than a less important financial decision, addressing an alternative explanation regarding decision importance. Finally, Study 5 found that even when the AI acknowledged the difficulty of making a moral tradeoff regardless of speed, which leveled up anthropomorphism differences, the effect persisted, corroborating the role of moral intuitions as a primary mechanism.

Table 1. Main results from all studies. Slower moral tradeoffs by AI improved impression of the AI in all studies except in conditions where relevant moral intuitions were absent or overridden (in grey). Means are followed by 95% confidence intervals.

<table>
<tr><th>Study</th><th>N</th><th>Context</th><th colspan="4">Moral-Faster Condition</th><th colspan="4">Same-Time Condition</th><th colspan="4">Moral-Slower Condition</th><th>Omnibus Test Results</th></tr>
<tr><td rowspan="2">Study 1</td><td rowspan="2">480</td><td rowspan="2">Trolley problems</td><td colspan="12">Impression of the AI (-5 = very negative, 5 = very positive)</td><td rowspan="2">$F(2, 477) = 56.05$, $p < .001$, $\eta_p^2 = .19$</td></tr>
<tr><td colspan="4">-.66
[-1.05, -.27]</td><td colspan="4">-.42
[-.70, -.14]</td><td colspan="4">1.65
[1.32, 1.97]</td></tr>
<tr><td>Study 2</td><td>375</td><td>Resource allocation problems</td><td colspan="4">-.89
[-1.35, -.43]</td><td colspan="4">.41
[-.05, .86]</td><td colspan="4">1.25
[.86, 1.64]</td><td>$F(2, 372) = 23.95$, $p < .001$, $\eta_p^2 = .11$</td></tr>
<tr><td rowspan="2">Study 3</td><td rowspan="2">401</td><td rowspan="2">Resource allocation problems</td><td colspan="12">% choosing a slower AI over a faster AI</td><td rowspan="2">$\chi^2(3) = 59.69$, $p < .001$, $\varphi = .22$</td></tr>
<tr><td colspan="3">Baseline (Average): 29.0%</td><td colspan="3">Moral-Duty (Average): 28.4%</td><td colspan="3">Non-moral Tradeoff (Average): 41.9%</td><td colspan="3">Moral-Tradeoff (Average): 54.9%</td></tr>
<tr><td>Study 4</td><td>300</td><td>Financial decisions</td><td colspan="4">Important-Faster: 1.71 [1.30, 2.12]</td><td colspan="4">Same-Time: 1.84 [1.49, 2.20]</td><td colspan="4">Important-Slower: 1.93 [1.55, 2.32]</td><td>$F(2, 297) = .32$, $p = .73$</td></tr>
<tr><td>Study 5</td><td>298</td><td>Trolley problems</td><td colspan="4">.48
[-.06, 1.02]</td><td colspan="4">1.31
[.83, 1.79]</td><td colspan="4">2.21
[1.81, 2.61]</td><td>$F(2, 295) = 13.00$, $p < .001$, $\eta_p^2 = .08$</td></tr>
</table>

Study 1: Trolley Problems

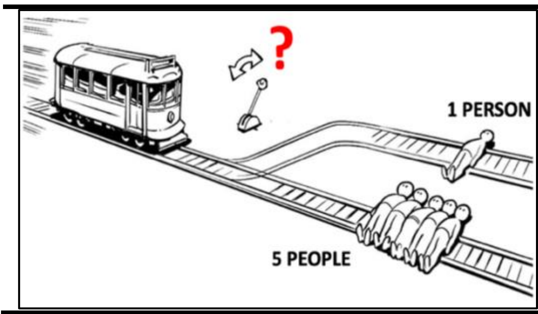
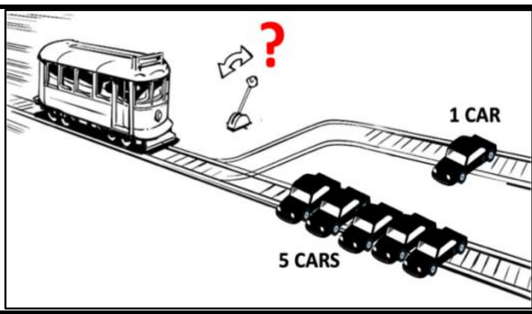
We pre-registered this study at https://aspredicted.org/3VD_ZXY.

Method

We requested 470 participants to a study link from a subject pool at a large public university for course credits. As pre-registered, we excluded all incomplete responses after one week of posting (see all screening details in Table S1, SOM). This resulted in 480 valid participants (M_{age} = 21, 264 female, 213 male, 3 non-binary/prefer not to say), all of whom received course credits.

Participants were asked to watch a short video about a state-of-the-art decision-making AI named NeuraYCG. In the video, NeuraYCG made two decisions, both illustrated (see Table 2) and explained. The first was the classic trolley dilemma (Foot, 1978): *"A runaway trolley was hurtling down the tracks toward five people, and the AI had to decide whether or not to press a lever to divert the trolley onto a different track and let it hit another person."* The second was structurally similar but without moral consequences: *"A runaway* *trolley was hurtling down the tracks toward five empty cars, and the AI had to decide whether or not to press a lever to divert the trolley onto a different track and let it hit another empty car*." A pretest (N = 149) validated that the two decisions were perceived to differ in moral difficulty (M_{first} = 3.95, 95% CI = [3.39, 4.50] vs. M_{second} = 3.14, 95% CI = [2.62, 3.65], $t(147)$ = 2.09, p = .038, Cohen's d = 0.34).

Table 2. Two trolley problems in Study 1, illustrated below and shown in the video.

Moral Tradeoff:	**Non-moral Decision:**
	

Participants were randomly assigned to one of three between-participants conditions (Moral-Faster, Same-Time, Moral-Slower). NeuraYCG's speed for each decision was displayed in the video in real time. In the Moral-Faster condition, NeuraYCG took 0.56 seconds to make the moral decision and 20.56 seconds to make the non-moral decision. In the Same-Time condition, NeuraYCG spent 10.56 seconds on both decisions. In the Moral-Slower condition, the moral decision took 20.56 seconds and the non-moral decision took 0.56 seconds. Its decision outcome was never revealed. In a manipulation check near the end of the study, over 80% of participants correctly recognized the speed variations in their assigned condition (in a multiple-choice question about if the AI took longer, the same time, or shorter *"to reach a decision for the life-loss decision than for the car-destruction decision"*; see details in SOM).

Participants were asked to rate their impression of the AI on an 11-point scale (-5 = very negative, 0 = neutral, 5 = very positive). Then, participants were asked to indicate the extent to which they agreed with the following statements on 7-point Likert scales (1 = strongly disagree, 7 = strongly agree). The first three items were on moral intuitions about the decision process, which we adapted from prior theorizing in Cushman (2008) and Tetlock et al. (2000): *"When deciding on potential life losses, NeuraYCG appears to show respect for the solemnity of the decision," "When deciding on potential life losses, I think the way that NeuraYCG makes this decision is morally objectionable (reverse-coded),"* and *"Irrespective of the decision outcomes, I feel that NeuraYCG makes decisions in a fair way."* On the next page were three items on anthropomorphism of the AI (Kim & McGill, 2011): *"I think NeuraYCG seems like a person," "...seems as if it has free will,"* and *"... seems as if it has intentions."* On the last page were four items on mind perception (Gray et al., 2011): *"I think NeuraYCG has the capacity to experience pleasure," "... experience desire," "... plan actions,"* and *"... exercise self-control."*

After those, participants were asked to indicate whether they thought NeuraYCG would press the lever in the moral decision. Finally, all participants were asked to briefly explain their impression ratings in an open-ended text box before filling out gender and age.

Results

Impression. Participants' impression of NeuraYCG differed by condition ($F(2, 477) = 56.05$, $p < .001$, $\eta_p^2 = .19$; Table 1 & Figure 1). Participants in the Moral-Slower condition ($M_{\text{Moral-Slower}} = 1.65$, 95% CI = [1.32, 1.97]) rated the AI more positively than did participants in the Moral-Faster condition (vs. $M_{\text{Moral-Faster}} = -.66$, 95% CI = [-1.05, -.27], pairwise $t(320) = 8.95$, $p < .001$, $d = 1.00$) or in the Same-Time condition (vs. $M_{\text{Same-Time}} = -.42$, 95% CI = [-.70, -.14], pairwise $t(320) = 9.44$, $p < .001$, $d = 1.05$). Impression did not differ between Moral-Faster and Same-Time conditions (pairwise $t(314) = .98$, $p = .33$).

Expected Outcome. Most participants predicted that NeuraYCG would press the lever. These predictions were similar across the three conditions (Moral-Faster 88.0% vs. Same-Time 92.4% vs. Moral-Slower 87.2%; $\chi^2(2, N = 480) = 2.61$, $p = .27$), and thus cannot explain impression differences. Among the participants with the "press" prediction, the speed effect remained significant and similar in size ($F(2, 425) = 57.31$, $p < .001$, $\eta_p^2 = .21$). Therefore, the impression differences were not driven by a small subset of participants with the "not press" prediction.

Moral Intuitions. Moral intuitions about the decision procedure also differed by condition (Cronbach's $\alpha = .20$[1]; $F(2, 477) = 42.41$, $p < .001$, $\eta_p^2 = .15$; Figure 1). Participants in the Moral-Slower condition had more positive moral intuitions regarding the decision procedure ($M_{\text{Moral-Slower}} = 4.41$, 95% CI = [4.30, 4.52]) than the other two conditions (vs. $M_{\text{Moral-Faster}} = 3.64$, 95% CI = [3.50, 3.79]; pairwise $t(320) = 8.39$, $p < .001$, $d = 0.94$; vs. $M_{\text{Same-Time}} = 3.84$, 95% CI = [3.73, 3.94]; pairwise $t(320) = 7.39$, $p < .001$, $d = 0.82$). Moral intuitions were also more positive in the Same-Time condition than in the Moral-Faster condition (pairwise $t(314) = 2.09$, $p = .037$, $d = 0.24$).

[1] This internal reliability score in Study 1 was lower than expected. Further analyses found that the reverse-worded second item, which should be negatively correlated with the other two items, was poorly correlated with the first item ($r = .03$, p =.47) and positively correlated with the third item ($r = .13$, $p = .006$) while the first and third items were moderately correlated ($r = .37$, $p < .001$). This score was marked higher ($\alpha = .63$) in Study 3, which used identical items, different decision scenarios, and Prolific participants, with all three items significantly correlated in the expected direction ($|r|$s > .14, ps < .006). Finally, in Study 6, the score further improved ($\alpha = .89$) using a reworded second item, trolley problems, and Prolific participants, with all three items strongly and positively correlated (rs > .68, ps < .001). Therefore, we speculated that the wording in the second item in Study 1were neglected or misread by the subject-pool participants, who are presumably less experienced with reverse wording in questionnaires. Despite the low Cronbach's α scores in some studies, the means of the collapsed items measuring moral intuitions followed a highly consistent pattern across all studies whenever measured (see also Studies S2&S3 in the SOM).

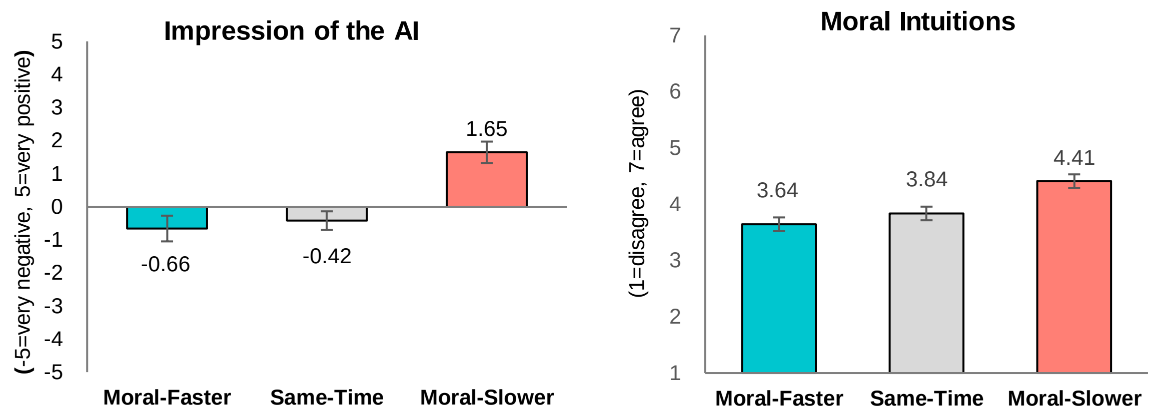


Figure 1. Impression of the AI (left) and moral intuitions (right) were the most positive in the Moral-Slower condition (in which the AI took longer to resolve a moral tradeoff than an otherwise comparable non-moral decision) in Study 1. All error bars indicate 95% confidence intervals.

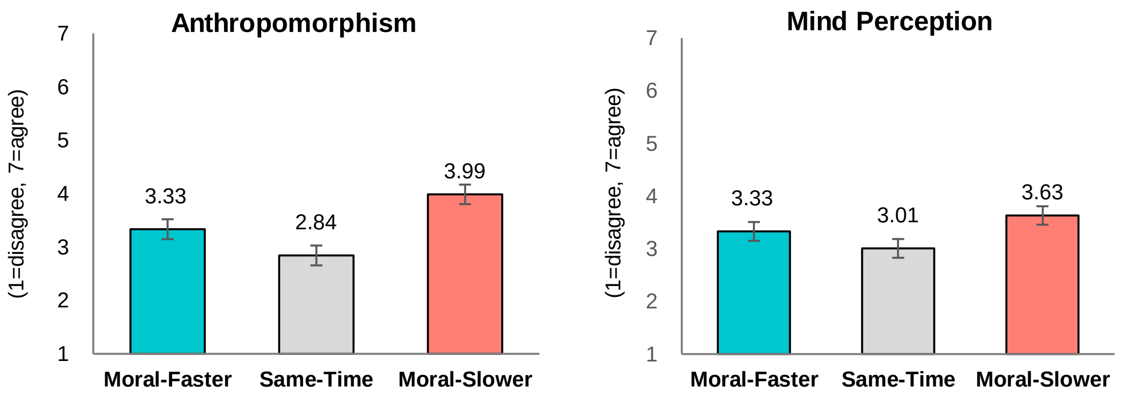


Figure 2. Results on anthropomorphism (left) and mind perception (right) in Study 1.

Anthropomorphism. Anthropomorphism (α = .78) differed by condition as well ($F(2, 477) = 37.28$, $p < .001$, $\eta_p^2 = .14$; Figure 2). Participants in the Moral-Faster condition ($M_{\text{Moral-Faster}} = 3.33$, 95% CI = [3.14, 3.52]) rated NeuraYCG as more human-like than in the Same-Time condition (vs. $M_{\text{Same-Time}} = 2.84$, 95% CI = [2.65, 3.04]; pairwise $t(314) = 3.55$, $p < .001$, $d = 0.40$). Participants in the Moral-Slower condition ($M_{\text{Moral-Slower}} = 3.99$, 95% CI = [3.81, 4.16]) rated NeuraYCG as more human-like than the other two conditions (pairwise ts > 5.01, ps < .001, ds > 0.55). Mind perception measures (α = .81; Figure 2) were highly similar to, albeit slightly weaker than anthropomorphism measures in all studies when measured. Therefore, we report the mind perception results in the SOM.

Mediations. Mediation analyses were conducted following the pre-registered criteria of observing meaningful differences on any process variables. First, in single mediation models, moral intuitions mediated the effect of speed variation (-1 = Moral-Faster, 0 = Same-Time, 1 = Moral-Slower) on AI impression (*indirect effect* = .56, *SE* = .07, 95% CI = [.43, .71]). Second, anthropomorphism also mediated this effect (*indirect effect* = .22, *SE* = .05, 95% CI = [.13, .33]), and so did mind perception (*indirect effect* = .07, *SE* = .03, 95% CI = [.01, .15]). In a parallel mediation model with all three mediator candidates, moral intuitions remained a robust mediator with the largest indirect effect (= .48, *SE* = .07, 95% CI = [.35, .61]). Anthropomorphism contributed with a smaller indirect effect (= .13, *SE* = .04, 95% CI = [.06, .22]), whereas mind perception no longer contributed (*indirect effect* = .003, *SE* = .02, 95% CI = [-.03, .04]).

Discussion

Study 1 presents initial evidence for the proposed effect: AI that takes longer to resolve a moral tradeoff is evaluated more favorably. Notably, participants' expectations regarding the AI's decision outcome on the moral tradeoff were similar across conditions, indicating that the observed effect is not driven by a belief that slower AI makes better decisions.

The process measures offer insights into the underlying mechanisms. First, participants in the Moral-Slower condition found the AI's way of resolving the moral tradeoff to be the most acceptable, and these intuitions mediated the observed effect. Second, while anthropomorphism also mediated the effect, albeit to a lesser extent, anthropomorphism alone does not fully explain the results. Participants liked the less-anthropomorphized AI in the

Same-Time condition more than the more-anthropomorphized AI in the Moral-Faster condition, suggesting that other factors are at play. Again, moral intuitions may fill this gap—when a human rapidly resolves a difficult moral tradeoff, it can be perceived as a sign of the person's irreverence or insolence, leading to negative moral intuitions that may also be overgeneralized onto AI in similar scenarios. This interpretation is supported by our pilot study, which was similar to Study 1 except that a person (instead of AI) made the decisions. The results showed a similar speed effect ($M_{\text{Moral-Faster}}$ = .54 vs. $M_{\text{Same-Time}}$ = .68 vs. $M_{\text{Moral-Slower}}$ = 1.73; $F(2, 300) = 8.25$, $p < .001$, $\eta_p^2 = .05$), also with similar expectations about the human decision-maker's choices (90.9% vs. 86.3% vs. 88.2%, $\chi^2(2, N = 303) = 1.06$, $p = .59$; S1, SOM). Additionally, unlike the AI, the human decision-maker was evaluated positively in all conditions (all Ms > 0), which aligns with the phenomena of AI aversion and human favoritism (Burton et al., 2020; Morewedge, 2022).

In summary, moral intuitions likely serve as the primary explanation for the observed effect, whereas anthropomorphism may play a secondary role.

Study 2: Resource Allocation Decisions

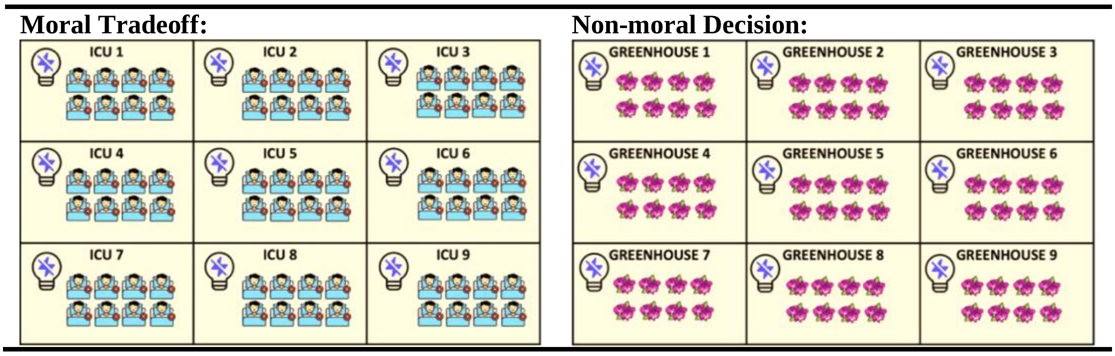
In Study 2, we sought to replicate the speed effect using a different type of moral tradeoff. We pre-registered the study at https://aspredicted.org/FCW_PYG.

Method

We recruited 375 participants (M_{age} = 42, 165 female, 200 male, 10 non-binary/prefer not to say) from Prolific US.

Participants read about an AI named NeuraYCG, which was tasked with two resource allocation decisions during a blizzard blackout (Table 3). In one, electricity was to be allocated to 2 out of 9 ICU units in a hospital; in the other, electricity was to be allocated to 2 out of 9 orchid greenhouses on a farm. Hence, the first was a moral tradeoff with ethical consequences, and the second had no ethical consequences. A pretest (N = 151) validated their differences in moral difficulty (M_{first} = 4.49, 95% CI = [3.93, 5.06] vs. M_{second} = 2.97, 95% CI = [2.53, 3.41]; $t(149) = 4.17$, $p < .001$, $d = 0.68$). Unlike the binary choices in the trolley problems, both decisions entailed multiple solutions.

Table 3. Two resource-allocation problems in Study 2.



Participants were randomly assigned to the three between-participants conditions (Moral-Faster, Same-Time, Moral-Slower). Unlike Study 1, the scenario and manipulations in Study 2 and subsequent studies were all described in text (without a video). Fast, average, and slow speeds were at 0.03, 5.23, and 10.43 seconds, respectively. More than 84% of all participants correctly recognized the speed differences near the end of the study (see SOM). All subsequent procedures were similar to Study 1, including measures of impression, moral intuitions, anthropomorphism, and mind perception.

Results

Impression. Impression differed by condition ($F(2, 372) = 23.95$, $p < .001$, $\eta_p^2 = .11$; Table 1). Participants in the Moral-Slower condition evaluated the AI more positively than did participants in the other two conditions (pairwise ts > 2.77, ps < .006, ds > 0.35).

Additionally, impression in the Same-Time condition was higher than in the Moral-Faster condition (pairwise $t(246) = 3.94$, $p < .001$, $d = 0.50$).

Moral Intuitions. Moral intuitions ($\alpha = .63$) also differed by condition ($F(2, 372) = 15.84$, $p < .001$, $\eta_p^2 = .08$). Participants had more positive moral intuitions in the Moral-Slower condition ($M_{\text{Moral-Slower}} = 4.31$, 95% CI = [4.11, 4.52]) than in the other two conditions (vs. $M_{\text{Moral-Faster}} = 3.47$, 95% CI = [3.26, 3.67]; pairwise $t(251) = 5.72$, $p < .001$, $d = 0.72$; vs. $M_{\text{Same-Time}} = 3.82$, 95% CI = [3.60, 4.04]; pairwise $t(247) = 3.21$, $p = .001$, $d = 0.41$). Additionally, moral intuitions were more positive in the Same-Time condition than in the Moral-Faster condition (pairwise $t(246) = 2.28$, $p = .023$, $d = 0.29$).

Anthropomorphism. Anthropomorphism ($\alpha = .88$) differed by condition as well ($F(2, 372) = 11.21$, $p < .001$, $\eta_p^2 = .06$). It was higher in the Moral-Slower condition ($M_{\text{Moral-Slower}} = 2.93$, 95% CI = [2.65, 3.22]) than in the other two conditions (vs. $M_{\text{Moral-Faster}} = 2.35$, 95% CI = [2.11, 2.59]; vs. $M_{\text{Same-Time}} = 2.08$, 95% CI = [1.85, 2.32]; pairwise ts > 3.09, ps < .002, ds > 0.38). It was directionally higher in the Moral-Faster condition than in the Same-Time condition (pairwise $t(246) = 1.53$, $p = .126$, $d = 0.20$). Mind perception results were similar to those on anthropomorphism (see SOM).

Mediations. In single mediation models, moral intuitions mediated the effect (*indirect effect* = .59, *SE* = .11, 95% CI = [.38, .81]); anthropomorphism (*indirect effect* = .17, *SE* = .06, 95% CI = [.06, .30]) and mind perception (*indirect effect* = .09, *SE* = .04, 95% CI = [.01, .18]) each mediated the effect with smaller indirect effects. In a parallel mediation model with all three mediator candidates, moral intuitions remained robust with the largest indirect effect (= .56, *SE* = .10, 95% CI = [.36, .77]), anthropomorphism contributed with a smaller indirect effect (= .07, *SE* = .03, 95% CI = [.01, .14]), and mind perception had no indirect effect (= -.009, *SE* = .02, 95% CI = [-.06, .03]).

Discussion

Study 2 extended the generalizability of the speed effect to resource-allocation decisions, moving beyond the conflict between deontological and utilitarian values in the stylized trolley dilemma. The process results were also replicated, supporting both moral intuitions and anthropomorphism as plausible mechanisms.

We have also replicated the effect in a variety of other contexts. Consistently, the focal moral tradeoff was paired with a non-moral decision of a similar structure that lacked implications for human life outcomes. These include other versions of trolley problems (S2, S4, & S8), tragic tradeoffs (S3 & S5), the autonomous vehicle dilemma (S6), and search decisions (S7). As Table 4 shows, these consistent results corroborate the effect's robustness to different decision parameters and situational factors including outcome scope, deliberation, urgency, life vividness, and specified decision outcomes. These supplemental studies also yielded similar process evidence to those reported in the main studies (see SOM).

Table 4. Main results from all supplemental studies.

Study	*N*	Context	Moral-Faster Condition	Same-Time Condition	Moral-Slower Condition	Additional Variable	Omnibus Test Results
S1	303	Trolley problems	.54 [.06, 1.01]	.68 [.25, 1.10]	1.73 [1.29, 2.16]	Human decision-makers	$F(2, 300) = 8.25$, $p < .001$, $\eta_p^2 = .05$
S2	299	Trolley problems	.19 [-.29, .67]	.77 [.31, 1.23]	1.91 [1.45, 2.37]	--	$F(2, 296) = 13.55$, $p < .001$, $\eta_p^2 = .08$
S3	305	Resource allocation	-.42 [-.94, .10]	-.01 [-.52, .50]	.73 [.22, 1.24]	--	$F(2, 302) = 4.92$, $p = .008$, $\eta_p^2 = .03$
S4	901	Trolley problems	.64 [.34, .93]	1.13 [.89, 1.37]	1.79 [1.53, 2.05]	Robust to scope	$F(2, 895) = 18.31$, $p < .001$, $\eta_p^2 = .04$
S5	1,205	Resource allocation	.24 [-.01, .49]	.40 [.14, .66]	.77 [.52, 1.02]	Robust to deliberation	$F(2, 1199) = 4.54$, $p = .011$, $\eta_p^2 = .01$
S6	1,199	Autonomous vehicle dilemmas	-.79 [-1.05, -.52]	.40 [.14, .65]	.45 [.18, .73]	Robust to urgency	$F(2, 1193) = 25.66$, $p < .001$, $\eta_p^2 = .04$
S7	1,203	Search problems	1.87 [1.67, 2.07]	2.06 [1.86, 2.26]	2.32 [2.12, 2.52]	Robust to life vividness	$F(2, 1197) = 4.82$, $p = .008$, $\eta_p^2 = .01$
S8	1,204	Trolley problems	.53 [.25, .81]	.85 [.58, 1.13]	.98 [.70, 1.25]	Robust to outcomes	$F(2, 1198) = 3.00$, $p = .050$, $\eta_p^2 = .01$

Study 3: Will People Choose a Slower AI?

The purpose of Study 3 was three-fold. First, we turned to a choice paradigm to test if and when people would prefer a generally slower AI over a generally faster AI. Moreover, to examine "moral" and "tradeoff" as two necessary conditions to the effect, we introduced conditions in which the slow AI decisions were non-moral tradeoffs (removing "moral") and moral-duty decisions (removing "tradeoff"), respectively. We pre-registered the study at https://aspredicted.org/RDS_Q7Y.

Method

We recruited 401 participants (M_{age} = 38, 162 female, 230 male, 9 non-binary/prefer not to say) from Prolific US.

Participants read about two decision-making AI models, both of which had been tested on thousands of decision tasks. Model A typically takes 0.01 – 0.30 seconds to make a decision, whereas Model B typically takes 0.06 – 0.70 seconds to do so. An attention check confirmed that over 95% of all participants identified Model B as generally slower than Model A.

Participants were randomly assigned to one of four (Baseline vs. Non-Moral Tradeoff vs. Moral Duty vs. Moral Tradeoff) between-participants conditions. In the Baseline condition, participants were not given further information. In the other conditions, participants read that Model A took 0.01 seconds to reach a specific decision, while *"Model B took 3.84 seconds, significantly longer than its typical response time."* Table 5 lists the specific problem corresponding to each condition.

Table 5. AI Model B was extra slow on a specific problem, corresponding to each condition.

Condition	Highlighted Decision Problem
Baseline	None.
Non-Moral Tradeoff	*"A company spends part of its annual profit on rewarding its star employees. This year, they decided to buy a new sports car for their CEO. The company needs to choose a good model among a wide variety of sports cars available on the market that has satisfying performance and is under the budget."*
Moral Duty	*"A company allocates part of its annual profit on social expenditures. This year, they can use it to support children's education in a rural village so that five children facing financial hardship can continue schooling. Or the company can spend it on buying a new sports car for the company's CEO."*
Moral Tradeoff	*"A company dedicates part of its annual profit to support children's education in rural areas. This year, a total of 87 primary school-aged youngsters in a rural village need financial assistance for their schooling. Unfortunately, the company's charity fund only has enough funding to support 5 children. These children are from different families and face different hardships."*

As our dependent variables, participants were asked to choose between the two models in response to the following questions: *"Which AI model do you prefer to use for decision-making?", "Which AI do you trust more?",* and *"Suppose your government is considering adopting one of the AIs for decision-making, which one would you support?"*

Next, moral intuitions were measured in pairs (*"I feel that Model A/Model B treats morally relevant decisions appropriately," "I feel the way that Model A/Model B makes morally relevant decisions is morally acceptable," and "I feel the way that Model A/Model B makes decisions is fair"*; 1 = strongly disagree, 7 = strongly agree). As pre-registered, we subtracted the ratings of Model A from those of Model B, and collapsed the difference scores. The resulting moral-intuition index ranged from -6 to 6, where a higher score indicates more positive moral intuitions regarding Model B relative to Model A.

In the Moral-Duty condition, participants were additionally asked to predict each model's decision (between "*buy a new sports car for the CEO*" and "*support five children's education*"). Most participants predicted that both AI models would similarly choose the latter (81.4% and 83.3%, p = .73). Expected decisions in Non-Moral-Tradeoff and Moral-Tradeoff conditions were unspecifiable, thus not measured.

Results

Choices. Most participants chose Model A in all choices except those in the Moral-Tradeoff condition (Table 6). Logistic regressions with a contrast dummy D1 (1, 1, 1, vs. -3 for Baseline, Moral Duty, Non-Moral Tradeoff, vs. Moral Tradeoff) showed that choices in the Moral-Tradeoff condition differed significantly from those in the other three conditions.

Table 6. Percentages of participants choosing the faster AI (Model A) in Study 3.

Choices	**Baseline (D1 = 1)**	**Moral Duty (D1 = 1)**	**Non-Moral Tradeoff (D1 = 1)**	**Moral Tradeoff (D1 = -3)**	**Omnibus Test Results**	**Logistic Regression Results for D1**
"Prefer"	80.8%	75.5%	64.4%	**47.5%**	$\chi^2(3, N = 401) = 29.29$, $p < .001$, $\varphi = .27$	$b = -.29$, $SE = .06$, Wald = 22.72, $p < .001$
"Trust"	66.7%	70.6%	54.5%	**44.4%**	$\chi^2(3, N = 401) = 17.61$, $p = .001$, $\varphi = .21$	$b = -.20$, $SE = .06$, Wald = 11.66, $p = .001$
"Support"	65.7%	68.6%	55.4%	**43.4%**	$\chi^2(3, N = 401) = 16.02$, $p = .001$, $\varphi = .20$	$b = -.20$, $SE = .06$, Wald = 11.92, $p = .001$

Moral intuition. Participants in the Moral-Tradeoff condition ($M_{\text{Moral Tradeoff}} = .65$, 95% CI = [.37, .93]) had the most positive moral intuitions for the slower AI ($\alpha = .93$), more than any other conditions (vs. $M_{\text{Baseline}} = .19$, 95% CI = [.06, .32] vs. $M_{\text{Non-moral Tradeoff}} = .37$, 95% CI = [.16, .59] vs. $M_{\text{Moral Duty}} = -.02$, 95% CI = [-.32, .29]; $F(3, 397) = 5.12$, $p = .002$, $\eta_p^2 = .04$; planned contrast $t(397) = -2.94$, $p = .004$ with weights 1, 1, 1 vs. -3 for Baseline, Non-Moral-Tradeoff, Moral-Duty, vs. Moral-Tradeoff conditions). The moral intuition index mediated the effect of decision type (with 1 = Moral-Tradeoff vs. 0 = any other condition) on all three choices (indirect effect on preference = .91, $SE = .34$, 95% CI = [.30, 1.66]; indirect effect on trust = 1.51, $SE = .57$, 95% CI = [.52, 2.77]; indirect effect on support for government adoption = 1.00, $SE = .39$, 95% CI = [.33, 1.86]).

Discussion

Participants chose a slower AI over a faster AI—only when the slower AI was particularly slow at resolving a moral tradeoff. This effect did not emerge when the AI was slow at resolving non-moral tradeoffs or moral-duty decisions. These results support our proposals that both “moral” and “tradeoff” are necessary for the effect to occur.

Study 4: Moral Relevance, or Outcome Importance?

One might wonder if the observed preference for the “Moral-Slower” AI is due to the greater importance of moral decisions than non-moral decisions. That is, perhaps people will like any AI that allocates more computational resources to the more important decision, moral or non-moral (i.e., the “cost-efficient AI” account). To address this alternative explanation, we tested the effect with a pair of more versus less important financial decisions in Study 5. We pre-registered this study at https://aspredicted.org/2C5_PK5.

Method

We recruited 300 participants (M_{age} = 29, 139 female, 156 male, 5 non-binary/prefer not to say) from Prolific US.

Participants read that a decision-making AI named BATY was used in two investment decisions. For the high-importance decision, BATY was asked to choose between two funds for a private equity firm with millions of dollars at stake: *“Fund A has a 92.7% chance of making a $5 million profit, and a 7.3% chance of incurring a $2 million loss. Fund B has a 5.5% chance of making a $10 million profit, and a 94.5% chance of incurring a $2 million loss.”* For the low-importance decision, BATY was asked to choose between two similar funds for a student club in an investment simulation game with millions of virtual tokens at stake that have no real-life value: “*Fund X has a 92.7% chance of earning 5 million virtual* *tokens, and a 7.3% chance of losing 2 million virtual tokens. Fund Y has a 5.5% chance of earning 10 million virtual tokens, and a 94.5% chance of losing 2 million virtual tokens.*"

A pre-test (N = 150) validated that the first decision was perceived as more important than the second one (important: 5.96 vs. 4.70; consequential: 5.50 vs. 4.61; financially costly: 6.14 vs. 4.50; ts > 3.53, ps < .001, ds > 0.57), while both were perceived as similar in moral difficulty (3.29 vs. 3.03, $t(148) = .96$, $p = .34$, $d = 0.16$).[2]

Participants were randomly assigned to one of three between-participants conditions (Important-Faster, Same-Time, Important-Slower). The fast, average, and slow speeds were at 0.0026, 0.6416, and 1.2806 seconds, respectively. Most participants in each condition (77.3%, 80.4%, and 76.2%) correctly identified the speed differences in a manipulation check. Participants were asked to rate their impression of the AI on an 11-point scale (-5 = very negative, 0 = neutral, 5 = very positive). Then, participants were asked to predict the AI's investment decisions before rating each decision on its importance, moral difficulty, and computational complexity. These ratings confirmed the results from the pre-test, where participants perceived the two decisions to differ only on importance, not on moral difficulty or computational complexity (see details in SOM).

Last, participants completed two attention checks *("Are there real financial consequences if the private equity firm's investment decision is unsuccessful?", "Are there real financial consequences if the student investment club's investment decision is unsuccessful?")*. Most participants (93.7%) indicated yes to the first attention check, and the majority (60.3%) indicated no to the second attention check. Although many participants

[2] For comparison, in the trolley problems (e.g., Studies 1 and 2) and the resource-allocation problems (e.g., Study 3), the first decisions (focal moral tradeoffs) were consistently perceived to be more morally difficult *and* more important than the second (non-moral) decisions (ps< .001). They were not perceived to consistently differ in computational complexity; see details in the pretests in SOM).

failed the second attention check, the main results held after excluding participants who failed at least one attention check (see SOM).

Results

Impression. Impression of the AI did not differ by condition ($F(2, 297) = .32$, $p = .73$; all pairwise *t*s < .77, *p*s > .44; Table 1 & Figure 3).

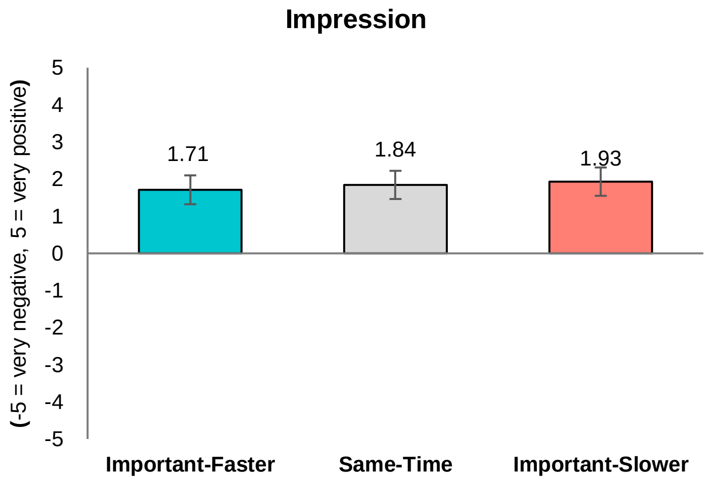


Figure 3. Impression of the AI were similar across conditions in Study 4.

Expected Outcome. On the high-importance decision, most participants expected the AI to choose Fund A, the more profitable option (Important-Faster: 97.9%; Same-Time: 94.1%; Important-Slower: 94.1%) at similar rates across conditions ($\chi^2(2, N = 300) = 2.19$, $p = .34$). On the low-importance decision, most participants expected the AI to choose Fund X, the more profitable option (90.7% vs. 88.2% vs. 78.2%), and these rates differed by condition ($\chi^2(2, N = 300) = 7.13$, $p = .028$, $\varphi = .15$). In other words, participants expected that a slower AI speed might decrease profits on the less important investment decision. However, these differences did not meaningfully affect impression of the AI, which was almost flat across conditions.

Discussion

AI's speed variations did not affect impression of the AI in Study 4 when the focal decision no longer involved a moral tradeoff. These results challenge the "cost-efficient AI" account. Therefore, in previous studies, the Moral-Slower AI was favored not merely for allocating processing resources in proportion to outcome importance.

However, is this null effect an artifact due to insufficient statistical power? Suppose a "true" effect went undetected in Study 4 due to inadequate sample size, based on the current effect size (d = 0.09), power analyses indicate that a total sample of 7,710 participants is required to detect it at 95% power with α = 0.05, or at least a total sample of 4,812 participants at 80% power. These targets are about 26 times and 16 times our current sample size (N = 300). They contrast with the results from other studies (e.g., Studies 3, 6, S2, and S3) with similar cell sizes, all of which showed a sizable effect (ds $\in$ [0.36, 0.70]). Therefore, we believe this null effect is unlikely an artifact of insufficient power. Further, we pooled the data from Study 5 and all comparable 3-cell studies (Studies 3, 6, S2, and S3) to explore whether the (moral vs. financial) decision type moderates the speed effect. This analysis yielded a significant 2-way interaction term ($F(2, 1571) = 7.68$, $p < .001$, $\eta_p^2 = .01$). Moreover, Bayesian analyses (Wagenmakers et al., 2018) revealed a very large Bayes factor, BF_{10} = 43.87, strongly supporting an interaction (over a null interaction) between speed variations and decision type on impression.

In summary, we find the null effect in Study 4 informative in addressing the alternative explanation related to outcome importance. These results also corroborate the moral relevance of the focal decision as necessary for the effect as shown in Study 3.

Study 5: Moral Intuitions vs Anthropomorphism

Finally, we tested between moral intutions versus anthropomorphism as two underlying mechanisms of the observed effect. We introduced an AI that acknowledged the difficulty of resolving moral tradeoffs regardless of its speed. We expect that this acknowledgement could level up anthropomorphism across conditions. That is, by acknowledging moral difficulty, the AI should appear more thoughtful and human-like, particularly in the Moral-Faster and Same-Time conditions where it might otherwise appear unthoughtful. Then, if anthropomorphism is the primary mechanism, the effect should be mitigated.

However, if moral intuitions are the primary mechanism, the speed effect should remain because moral intuitions are particularly sensitive to action cues (Cushman, 2013) such as decision speed, while remarkably resistant to justification and persuasion attempts (Haidt et a. 2000). This study was pre-registered at https://aspredicted.org/P6K_SM6.

Method

We recruited 298 participants (M_{age} = 42, 158 female, 134 male, 6 non-binary/prefer not to say) from Prolific US.

The stimuli were adapted from Study 1,but using a text-based manipulation (instead of video-based). In addition, NeuraYCG always returned the following message when working on the moral tradeoff: *"I find this decision very difficult to make and believe it should be addressed carefully. I made my decision based on my best judgment."* All measures were the same as in Study 1, except that the reverse-coded item in moral intuition endorsement was replaced with *"When deciding on potential life losses, I think the way that NeuraYCG makes this decision is morally acceptable"* to mitigate inattentive responses. All other process measures remained unchanged. The majority of participants recognized that NeuraYCG explicitly acknowledged moral difficulty (Moral-Faster: 53.5%; Same-Time:

59.8%; Moral-Slower: 69.0%) in an attention check near the end of the study, with marginally different rates across conditions ($\chi^2(2, N = 298) = 5.14$, $p = .077$, $\varphi = .13$). Despite lower recall rates than expected, all main results hold after excluding participants who failed the recall (see SOM).

Results

Impression. Impression differed by condition ($F(2, 295) = 13.00$, $p < .001$, $\eta_p^2 = .08$; Figure 4). Participants in the Moral-Slower condition evaluated the AI more positively than those in the Moral-Faster (pairwise $t(199) = 5.08$, $p < .001$, $d = 0.72$) and Same-Time conditions (pairwise $t(195) = 2.84$, $p = .005$, $d = 0.41$). Impression in the Same-Time condition was also higher than in the Moral-Faster condition (pairwise $t(196) = 2.28$, $p = .024$, $d = 0.32$).

Figure 4. In Study 5, impression and moral intuitions were still the most positive in the Moral-Slower condition, while anthropomorphism was similar across conditions ($p = .29$).

Moral Intuitions. Moral intuitions ($\alpha = .89$) similarly differed by condition ($F(2, 295) = 7.82$, $p < .001$, $\eta_p^2 = .05$; Figure 5). They were higher in the Moral-Slower condition ($M_{\text{Moral-Slower}} = 5.30$, 95% CI = [5.05, 5.55]) than in the other two conditions (vs. $M_{\text{Moral-Faster}} = 4.60$, 95% CI = [4.34, 4.85], pairwise $t(199) = 3.83$, $p < .001$, $d = 0.54$; vs. $M_{\text{Same-Time}} = 4.81$, 95% CI = [4.56, 5.06], pairwise $t(195) = 2.68$, $p = .008$, $d = 0.38$). They did not differ between the Moral-Faster and Same-Time conditions (pairwise $t(196) = 1.17$, $p = .243$).

Anthropomorphism. Anthropomorphism ($\alpha = .88$) no longer differed between any two conditions ($M_{\text{Moral-Faster}} = 3.26$, 95% CI = [2.96, 3.57] vs. $M_{\text{Same-Time}} = 3.54$, 95% CI = [3.20, 3.87] vs. $M_{\text{Moral-Slower}} = 3.61$, 95% CI = [3.28, 3.95]; $F(2, 295) = 1.25$, $p = .29$; pairwise ts < 1.52, ps > .132; Figure 4). Mind perception was also flat ($F(2, 295) = .91$, $p = .41$; pairwise ts < 1.27, ps > .209).

Mediation. Moral intuitions still mediated the effect (*indirect effect* = .42, *SE* = .12, 95% CI = [.20, .66]). Anthropomorphism did not (*indirect effect* = .13, *SE* = .09, 95% CI = [-.04, .31]), nor did mind perception.

Discussion

The AI's acknowledgment of moral difficulty led to two distinct outcomes. On the one hand, participants perceived the AI as similarly human-like across all conditions, regardless of speed variations. On the other hand, speed variations still caused differences in moral intuitions across conditions. Consequently, the speed effect persisted, supporting the primary role of action-induced moral intuitions in driving the observed effect.

Additional Mechanism Insights from Text Analyses with GPT-4o

We employed a GPT-4o to analyze all 6,265 entries of text responses that we collected from nine studies using the speed-variation paradigm, to obtain additional insights into potential psychological mechanisms. Importantly, text analyses are subject to limitations of introspection and verbalization (Nisbett & Wilson, 1977)

Identifying Target Mechanisms

First, we planned to code for all primary and secondary mechanisms measured in our studies. These include the three items of "moral intuitions," separately coded. We also included an item for anthropomorphism ("The AI seems like a person, has free will, or seems as if it has intentions") and an item for mind perception ("The AI has capacities commonly associated with a mind, such as experiencing pleasure, having desires, planning actions, or exercising self-control").

Second, we searched prior literature for factors influencing AI-related attitudes. To do so, we used an LLM (GPT-4o) to find all articles published in top-tier peer-reviewed journals (ABS journal level 4 or 4*) in psychology, marketing, and management from 1950 to 2023.[3] This search identified 297 relevant articles. Then we asked two research assistants, who were blind to our hypotheses, to identify psychological mechanisms studied in these articles that contribute to AI aversion. A total of 23 unique psychological mechanisms in 23 articles (e.g., people believe that AI cannot properly account for recipients' individual characteristics and thereby dislike its use; Longoni, Bonezzi & Morewedge 2019) were identified, besides the 5 items we measured in a subset of our studies. Table S3 in the SOM summarizes all 28 target mechanisms and a one-sentence summary of each mechanism that was used as the prompt for subsequent coding.

Response Coding

Next, we used GPT-4o to code each text response for the target mechanisms. An inclusive prompt was used for all target mechanisms, assigning a "1" if the response mentioned the mechanism in any form (e.g., that AI processes individual characteristics in decision-making differently than humans do), and "0" otherwise. For each mechanism, a short prompt and 15 pre-classified examples were provided to GPT-4o, following a few-shot

[3] The search included the following keywords: "Artificial intelligence OR AI OR algorithm OR service bots OR robots OR chatbots OR automation OR autonomous vehicles" AND "decision making OR decision aids OR evaluation OR attitude OR judgment."

learning approach that has been validated by prior studies (e.g., Loukas et al., 2023). We automated this process by using Chat GPT API.

These examples were randomly drawn from all responses, 5 from each condition (Moral-Faster vs. Same-Time vs. Moral-Slower), and pre-classified by two research assistants independently. Disagreements between coders (< 10%) were resolved through discussion. The full list of examples is provided on OSF.

Method Validation

We validated the coding results by GPT-4o by having two research assistants independently classify a subset of text entries using the same prompts and examples for 3 target mechanisms—the most frequently mentioned, a moderate-frequently mentioned, and the least frequently mentioned, as identified by GPT-4o. For each mechanism, one study was randomly selected out of five studies with a sample size of 500 or less. Two RAs coded each response independently, and any disagreements were resolved through discussion. The classified responses using both methods and correlations between methods are listed in the web appendix.

Results

Moral Intuitions. Overall, items related to the three moral-intuition items were detected 715 times across all responses, the most frequent among all. Thoughts related to the first item ("whether the AI showed respect for the solemnity of the decisions involving potential loss of life") were detected 413 times (6.59% of all): 56 times in the Moral-Faster condition, 28 times in the Same-Time condition, and 329 times in the Moral-Slower condition, ($F(2, 6262) = 224.76$, $p < .001$). Thoughts related to the second item ("whether the AI's decision-making in scenarios involving potential life loss is morally objectionable") were detected 298 times (4.76% of all), which did not vary significantly across conditions ($F(2, 6262) = .25$, $p = .78$). Thoughts related to the third item ("whether the AI's decision-making is fair, irrespective of the outcome") was detected only four times (.06% of all) and did not vary significantly across conditions (p = .77). Results were similar between studies in which these items were measured before the open-ended responses, and studies in which the items were not measured (see web appendix). Therefore, the frequent mention of "moral intuitions" in participants' open-ended responses cannot be solely explained by exposure to related items in the preceding questions. These results support the role of moral intuitions, especially those regarding a sense respect associated with the decision procedure (per the first item), underlying the effect. These results also suggest that thoughts related to a sense of respect were presumably more accessible than other aspects of speed-related moral intuitions, even though ratings differed on the measured items.

Anthropomorphism. Thoughts related to anthropomorphism and mind perception were mentioned relatively less frequently, detected 34 times (.54% of all) and 11 times (.18% of all), respectively. Additionally, when we coded responses for mentions of the AI displaying human-like attributes—a mechanism identified by de Visser (2016), one of the 23 mechanisms identified from prior literature—these mentions comprised 2.79% of total responses, suggesting that anthropomorphism does not primarily drive the observed effects. Nonetheless, we observed significant variation in frequency across conditions ($F(2, 6262) = 38.67$, $p < .001$). These mentions occurred in 1.49% in the Moral-Faster condition, 1.50% in the Same-Time condition, and 5.34% of responses in the Moral-Slower condition, consistent with the results of measured anthropomorphism.

Among the other 22 mechanisms identified in prior literature, only one appeared in more than 5% of all responses: the belief about the (algorithm or human) decision-maker's tendency to be influenced by individual characteristics, as identified by Bonezzi & Ostinelli (2021), was detected in 5.11% of responses. Thoughts related to this mechanism differed in frequency across conditions ($F(2, 6262) = 4.38$, $p = .010$), with 4.82%, 6.23%, and 4.28%

mentions in the Moral-Faster, Same-Time, and Moral-Slower conditions, respectively. However, these differences were not aligned with the "slower is better" effect and could not serve as a primary explanation.

General Discussion

Across 13 pre-registered experiments (*N* = 8,473), participants rated an AI more favorably when it takes longer to make moral-tradeoff decisions, even when the decision outcomes were identical. Both the "moral" and "tradeoff" elements of the focal decisions were necessary for the effect. Process measures indicate that overgeneralized moral intuitions primarily drove this preference, with anthropomorphism playing a secondary role. Further analyses of open-ended text responses using GPT-4o corroborated the process evidence.

Our findings caution researchers against casual assumptions that AI aversion arises *despite* people appreciating its fast speed. Instead, we show that AI's fast speed may itself contribute to AI aversion in moral contexts. As AI-based decision aids become increasingly integrated into policy decision systems, understanding the causes of AI aversion is crucial. This research offers fresh insights into the psychology of AI aversion in moral domains and highlights new avenues for mitigating AI aversion.

References

Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral Judgment and Decision Making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 478–515). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118468333.ch17

Bartels, D. M., & Medin, D. L. (2007). Are Morally Motivated Decision Makers Insensitive to the Consequences of Their Choices? *Psychological Science*, *18*(1), 24–28. https://doi.org/10.1111/j.1467-9280.2007.01843.x

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

Bonnefon, J.-F., & Rahwan, I. (2020). Machine Thinking, Fast and Slow. *Trends in Cognitive Sciences*, *24*(12), 1019–1027. https://doi.org/10.1016/j.tics.2020.09.007

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

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

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

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How Quick Decisions Illuminate Moral Character. *Social Psychological and Personality Science*, *4*(3), 308–315. https://doi.org/10.1177/1948550612457688

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366. https://doi.org/10.1016/j.tics.2013.06.005

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. https://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F. (2013). Action, Outcome, and Value: A Dual-System Framework for Morality. *Personality and Social Psychology Review*, *17*(3), 273–292. https://doi.org/10.1177/1088868313495594

Cushman, F., Young, L., & Greene, J. (2010). *Our multi-system moral psychology: Towards a consensus view*.

Cushman, F., Young, L., & Hauser, M. (2006). The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm. *Psychological Science*, *17*(12), 1082–1089. https://doi.org/10.1111/j.1467-9280.2006.01834.x

Dabrowski, J., & Munson, E. V. (2011). 40years of searching for the best computer system response time. *Interacting with Computers*, *23*(5), 555–564. https://doi.org/10.1016/j.intcom.2011.05.008

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents. *JOURNAL OF EXPERIMENTAL PSYCHOLOGY-APPLIED*, *22*(3), 331–349. https://doi.org/10.1037/xap0000092

Dietvorst, B. J., & Bartels, D. M. (2022). Consumers Object to Algorithms Making Morally Relevant Tradeoffs Because of Algorithms' Consequentialist Decision Strategies. *Journal of Consumer Psychology*, *32*(3), 406–424. https://doi.org/10.1002/jcpy.1266

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science, 31*(10), 1302–1314. https://doi.org/10.1177/0956797620948841

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

Efendić, E., Van de Calseyde, P. P. F. M., & Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational Behavior and Human Decision Processes*, *157*, 103–114. https://doi.org/10.1016/j.obhdp.2020.01.008

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Feiner, H. F., Lauren. (2023, October 30). *Biden issues U.S.' first AI executive order, requiring safety assessments, civil rights guidance, research on labor market impact*. CNBC. https://www.cnbc.com/2023/10/30/biden-unveils-us-governments-first-ever-ai-executive-order.html

Foot, P. (1978). The Problem of Abortion and the Doctrine of the Double Effect. In *Virtues and Vices* (pp. 19–32). Oxford: Blackwell.

Goodman, T., & Spence, R. (1978). The effect of System Response Time on interactive computer aided problem solving. *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*, 100–104. https://doi.org/10.1145/800248.807378

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of Mind Perception. *Science, 315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Gray, K., Knobe, J., Sheskin, M., Bloom, P., & Barrett, L. F. (2011). More than a body: Mind perception and the nature of objectification. *Journal of Personality and Social Psychology*, *101*(6), 1207–1220. https://doi.org/10.1037/a0025883

Gray, K., Waytz, A., & Young, L. (2012). The Moral Dyad: A Fundamental Template Unifying Moral Judgment. *Psychological Inquiry*, *23*(2), 206–215. https://doi.org/10.1080/1047840X.2012.686247

Greene, J. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323. https://doi.org/10.1016/j.tics.2007.06.004

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Sciences*, *6*(12), 517–523. https://doi.org/10.1016/S1364-6613(02)02011-9

Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haidt, J., Björklund, F., & Murphy, S. (2000). *Moral dumbfounding: When intuition finds no reason*. https://polpsy.ca/wp-content/uploads/2019/05/haidt.bjorklund.pdf

Hutchinson, A. (2023). How machine learning might unlock earthquake prediction. *MIT Technology Review*. https://www.technologyreview.com/2023/12/29/1084699/machine-learning-earthquake-prediction-ai-artificial-intelligence/

Jauernig, J., Uhl, M., & Walkowitz, G. (2022). People Prefer Moral Discretion to Algorithms: Algorithm Aversion Beyond Intransparency. *Philosophy & Technology*, *35*(1), 2. https://doi.org/10.1007/s13347-021-00495-y

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004

Kahneman, D., & Frederick, S. (2005). A Model of Heuristic Judgment. In *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge University Press.

Kahneman, D., & Sunstein, C. (2007). *Indignation: Psychology, Politics, Law Indignation: Psychology, Politics, Law*. 12697–12706.

Kim, S., & McGill, A. L. (2011). Gaming with Mr. Slot or Gaming the Slot Machine? Power, Anthropomorphism, and Risk Perception. *Journal of Consumer Research*, *38*(1), 94–107. https://doi.org/10.1086/658148

Kohlberg, L. (1969). *Stage and sequence: The cognitive-developmental approach to socialization* (p. 134). Garland Publishing.

Kupor, D. M., Tormala, Z. L., Norton, M. I., & Rucker, D. D. (2014). Thought Calibration: How Thinking Just the Right Amount Increases One’s Influence and Appeal. *Social Psychological and Personality Science*, *5*(3), 263–270. https://doi.org/10.1177/1948550613499940

Landy, J. F., Herzog, N. R., & Bartels, D. M. (2024). *Thoughtfulness is a Virtue: Evaluations of Decision-Makers in Complex Moral Situations*.

Mankowitz, D. J., Michi, A., Zhernov, A., Gelmi, M., Selvi, M., Paduraru, C., Leurent, E., Iqbal, S., Lespiau, J.-B., Ahern, A., Köppe, T., Millikin, K., Gaffney, S., Elster, S., Broshear, J., Gamble, C., Milan, K., Tung, R., Hwang, M., … Silver, D. (2023).

Faster sorting algorithms discovered using deep reinforcement learning. *Nature*, *618*(7964), 257–263. https://doi.org/10.1038/s41586-023-06004-9

Mehdi, Y. (2024, May 20). *Introducing Copilot+ PCs*. The Official Microsoft Blog. https://blogs.microsoft.com/blog/2024/05/20/introducing-copilot-pcs/

Morewedge, C. K. (2022). Preference for human, not algorithm aversion. *Trends in Cognitive Sciences*, *26*(10), Article 10. https://doi.org/10.1016/j.tics.2022.07.007

Morewedge, C., Naushan, H. F., & Mullainathan, S. (2024). Decisions With Algorithms. In *The Handbook of Social Psychology* (6th ed.). Cambridge University Press.

Morewedge, C., Preston, J., & Wegner, D. M. (2007). Timescale bias in the attribution of mind. *Journal of Personality and Social Psychology*, *93*(1), 1–11. https://doi.org/10.1037/0022-3514.93.1.1

Narayan, J., Hu, K., Coulter, M., Mukherjee, S., Hu, K., & Mukherjee, S. (2023, April 5). Elon Musk and others urge AI pause, citing “risks to society.” *Reuters*. https://www.reuters.com/technology/musk-experts-urge-pause-training-ai-systems-that-can-outperform-gpt-4-2023-03-29/

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231.

Nvidia. (2017). NVIDIA Announces World's First AI Computer to Make Robotaxis a Reality. NVIDIA Newsroom, nvidianews.nvidia.com/news/nvidia-announces-world-s-first-ai-computer-to-make-robotaxis-a-reality

OpenAI. (2024). Hello GPT-4o. *OpenAI*. https://openai.com/index/hello-gpt-4o/

Shneiderman, B. (1984). Response time and display rate in human performance with computers. *ACM Computing Surveys*, *16*(3), 265–285. https://doi.org/10.1145/2514.2517

Shurman, E., & Brutlag, J. (2009). *Performance related changes and their user impact – ScienceOpen*. Velocity 2009. https://www.scienceopen.com/document?vid=d2ce0ed5-d801-4779-8182-91c2875c82a2

Sinnott-Armstrong, W., Young, L., & Cushman, F. (2010). Moral Intuitions. In J. M. Doris & The Moral Psychology Research Group (Eds.), *The Moral Psychology Handbook* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199582143.003.0008

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., & Etzioni, O. (2016). *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel*. https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/ai100report10032016fnl_singles.pdf

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal*, *94*. https://doi.org/10.2307/796133

Valenzuela, A., Puntoni, S., Hoffman, D., Castelo, N., De Freitas, J., Dietvorst, B., Hildebrand, C., Huh, Y. E., Meyer, R., Sweeney, M. E., Talaifar, S., Tomaino, G., & Wertenbroch, K. (2024). How Artificial Intelligence Constrains the Human Experience. *Journal of the Association for Consumer Research*, *9*(3), 241–256. https://doi.org/10.1086/730709

Van de Calseyde, P. P. F. M., Keren, G., & Zeelenberg, M. (2014). Decision time as information in judgment and choice. *Organizational Behavior and Human Decision Processes*, *125*(2), 113–122. https://doi.org/10.1016/j.obhdp.2014.07.001

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Xie, Y., Qu, J., Zhang, Y., Zhou, R., & Chan, A. H. S. (2023). Speaking, fast or slow: How conversational agents' rate of speech influences user experience. *Universal Access in the Information Society*. https://doi.org/10.1007/s10209-023-01000-2

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*, 103870. https://doi.org/10.1016/j.jesp.2019.103870