European Economic Review 184 (2026) 105255

Contents lists available at ScienceDirect

European Economic Review

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

Trusting machines with morality — Delegating moral decisions to AI☆

Nicola Hüholt *, Nora Szech

Chair of Political Economy, Karlsruhe Institute of Technology, Fritz-Erler-Str. 1-3, Karlsruhe, 76133, Germany

ARTICLE INFO

Keywords:
Delegation
Artificial intelligence
Moral decision-making
Algorithm aversion

ABSTRACT

Research suggests that individuals are generally skeptical about the use of artificial intelligence (AI) in moral contexts, favoring human decision-makers over AI. Yet, in two experiments involving a total of 5639 participants, we find that individuals facing a real-life moral decision delegate significantly more often when they can delegate to AI rather than to a human counterpart. This result highlights AI's relative appeal as a moral delegate, indicating that individuals' preferences for AI's involvement change when they themselves assume the role of a decision-maker. Responsibility shifting, previously studied as a motive for delegation to humans, extends to AI delegates. Moreover, it appears to be facilitated by individuals adapting their beliefs about AI's capability in a self-serving manner. Ambiguity surrounding that capability allows them to interpret it in ways that justify delegation. These findings add nuance to assumptions about algorithm aversion in moral domains and raise critical questions about accountability and the ethical implications of relying on AI for morally sensitive decisions.

1. Introduction

Artificial intelligence (AI) has become an integral component across a wide range of industries, many of which intersect with ethical considerations (Bonnefon et al., 2024; Wallach and Allen, 2008). AI tools are already used to make or support decisions in healthcare about the allocation of limited resources (Obermeyer et al., 2019), in finance to determine mortgage or loan eligibility (Hale, 2021; Zou and Khern-am nuai, 2023), and in hiring processes to evaluate job candidates (Dattner et al., 2019; Dastin, 2022). In the most critical cases, they are tasked with making life-and-death decisions (Awad et al., 2018; Holbrook et al., 2024; Adam, 2024). These decisions, involving the distribution of well-being or harm among individuals, are often characterized by ethical trade-offs and therefore fall into the moral domain (Anderson and Anderson, 2011; Gert, 2005; Awad et al., 2018). While such real-world examples showcase AI's expanding role, they also reveal limitations — such as the replication of biases — and the need to account for AI-specific characteristics, such as the opacity of its decision-making (Gerke et al., 2020; Cath, 2018; Pazzanese, 2020). Given the sensitive nature and high stakes of these decisions, coupled with the growing availability of AI tools, questions arise about individuals' willingness to hand over responsibility to AI when faced with morally complex choices.

Our study addresses these questions, and demonstrates that delegation demand in a moral decision is significantly higher when the delegate option is an AI rather than a human. This finding nuances the prevailing notion that individuals generally feel *algorithm aversion* toward the use of AI in moral contexts (Castelo et al., 2019; Bigman and Gray, 2018), showing that aversion can even reverse into greater acceptance compared to a human counterpart. This increased demand to delegate to AI appears unaffected by

☆ This article is part of a Special issue entitled: 'Nora Szech' published in European Economic Review.

* Corresponding author.
E-mail address: nicola.hueholt@kit.edu (N. Hüholt).

https://doi.org/10.1016/j.euroecorev.2025.105255
Received 16 December 2024; Received in revised form 16 December 2025; Accepted 30 December 2025
Available online 8 January 2026

the severity of the moral dilemma, as we find higher delegation rates to AI irrespective of whether the moral decision is presented in a positive or negative decision context.

Preferences regarding AI's involvement are influenced by individuals' own role in the decision-making process. Unlike prior studies, which often rely on hypothetical dilemmas or scenarios, we employ a *real donation choice* inspired by structural elements of the trolley problem, thereby putting participants into the role of a *decision-maker* with real responsibility and a delegation option. In this setting, we examine if and how the mechanism of *responsibility shifting* — previously studied for delegation to humans (Bartling and Fischbacher, 2012) — extends to AI delegates. We find evidence that it does and may even be facilitated by AI's intrinsic features, contributing to the increased delegation demand. Opacity and ambiguity surrounding AI's capabilities for making moral decisions may create room for individuals to inflate their beliefs about the quality of AI's moral decision-making to justify transferring responsibility — akin to *moral wiggle room* (Dana et al., 2007) in the mechanistic sense of ambiguity-enabled self-justification. Accordingly, we observe that individuals rate AI's moral capabilities higher when they have the option to delegate to it, particularly when their decision has real consequences.

The results of this study highlight broader societal implications, emphasizing the need for conscious design and governance of AI systems in moral domains to preserve accountability and ethical standards.

Algorithm aversion

Our findings contrast with extant research which suggests that people are generally skeptical of AI's ability to handle moral decisions and want such decisions to remain under human control.

People can be reluctant to rely on algorithms or to allow them to make decisions even when the algorithm outperforms humans, a phenomenon termed *algorithm aversion* by Dietvorst et al. (2015). The term originally refers to deterministic, rule-based algorithms (Dietvorst et al. 2015; Castelo et al. 2019). Later research extends this focus to AI systems (Bigman and Gray, 2018; Jussupow et al., 2020). While some studies use the terms algorithm, AI, machines or automation interchangeably (e.g., Burton et al., 2020), others emphasize AI's unique attributes, such as its perceived mind and its moral reasoning capacity (e.g., Bigman and Gray, 2018; Zhang et al., 2022; Gogoll and Uhl, 2018). Many of the cognitive biases and concerns underlying algorithm aversion — such as distrust in computational decision-making and preference for human-like judgment — also apply to AI systems, though often with additional dimensions (Bigman and Gray, 2018; Zhang et al., 2022).

For decisions with moral aspects, algorithm aversion is especially pronounced (Chugunova and Sele, 2022; Mahmud et al., 2022; Jussupow et al., 2020). A key reason is that individuals perceive AI as lacking the capabilities required for moral decision-making. For instance, Bigman and Gray (2018) document widespread aversion to machines making moral decisions in paradigmatic moral dilemmas across various domains, including driving, legal, medical, and military contexts. Notably, the aversion persists even when machines' decisions lead to positive outcomes (Bigman and Gray, 2018). This aversion is attributed to the perception that machines lack a complete mind (*mind perception*) needed to make moral decisions (Bigman and Gray, 2018; Gray et al., 2012; Young and Monroe, 2019). *Mind perception* refers to the attribution of mental capacities, and has two dimensions: the ability to fully think (*agency*) and feel (*experience*) (Bigman and Gray, 2018; Gray et al., 2007; Waytz et al., 2010). Other researchers have found similar results in decisions involving subjective judgment, as well as in different morally sensitive scenarios like personal and impersonal high-stakes moral dilemmas, medical AI or consumer interactions (Castelo et al., 2019; Zhang et al., 2022; Longoni et al., 2019; Dietvorst and Bartels, 2022). Across these domains, concerns consistently stem from AI's perceived lack of essential (human) capabilities — including 'affective human-likeness' and warmth, sensitivity to individual nuances ('uniqueness neglect'), a tendency toward utilitarian or consequentialist reasoning, and insufficient intuition or subjective judgment capability. As a result, individuals perceive algorithmic decisions as less ethical and authentic, and therefore AI to not be suited for subjective tasks (Lee, 2018; Jago, 2017).

In line with such concerns, third-parties tend to judge delegation to machines more critically as well, with individuals rewarding the delegation choice less when the delegator selects a machine rather than a human (Gogoll and Uhl, 2018).

Ensuring transparency in decision-making and incorporating a 'human in the loop' as oversight and control mechanism can alleviate some of the aversion (Bigman and Gray, 2018). Measures such as these are also a core component of many legislative frameworks, which emphasize that sensitive moral decisions should ultimately remain in the hands of humans and are accessible to them (GDPR, 2016).

These insights into algorithm aversion suggest that people are critical of both the use of AI for moral decisions and its capability to make them, particularly in hypothetical or observer contexts or when personal stakes are involved. However, preferences can shift depending on context and the individual's role. For example, while people approve of autonomous vehicles programmed to sacrifice passengers to save others, they prefer not to ride in such vehicles themselves (Bonnefon et al., 2016). Analogously, when people act as moral decision-makers, they may accept AI more readily if delegation makes their lives easier. This reflects a highly relevant real-world scenario in which decision-makers have access to AI tools for making or supporting decisions with moral implications.

Delegation and shifting responsibility

Demand to delegate moral decisions has already been studied extensively for human delegates. Bartling and Fischbacher (2012) showed that some people prefer to delegate moral decisions to others instead of deciding themselves. A key factor in this behavior is a *shift of responsibility* attribution, both in individuals' own eyes and in the eyes of others. Delegation leads to more selfish behavior and at the same time reduces punishment from third parties (Bartling and Fischbacher, 2012; Coffman, 2011; Hamman et al., 2010),

and conversely, reduces rewards for positive or generous decisions (Argenton et al., 2023). Sharing or delegating the decision can reduce feelings of moral responsibility, guilt, or potential regret and help to keep a positive self-image (Bartling and Fischbacher, 2012; Rothenhäusler et al., 2018; Falk et al., 2020; Falk and Szech, 2013; Steffel and Williams, 2018; Bartling et al., 2023).

Delegating to create *moral wiggle room* and exploit ambiguity around the final moral outcome can also shift responsibility, protect self-image, and decrease punishment from others (Dana et al., 2007; Serra-Garcia and Szech, 2021; Bartling et al., 2014; Grossman and van der Weele, 2017). Fahrenwaldt et al. (2024) specify the mechanism behind moral wiggle room as situational features that hinder linking an agent's behavior clearly to self-serving intentions, leaving room for other justifications when behaving selfishly.

When facing a decision for others rather than oneself, self-serving behavior can also be motivated by psychological relief rather than financial reward. The desire to lower feelings of responsibility and regret impacts behavior even when it is not tied to material incentives: people are more likely to delegate then, especially when having to decide between two negative outcomes. In these contexts, the delegate's expertise is secondary — instead, what matters to decision-makers is that responsibility can be transferred (Steffel et al., 2016; Steffel and Williams, 2018). Because AI systems are often opaque and their competence is hard to verify, they may promote such dynamics.

However, while such motives and outcomes are well established for *human delegates*, their applicability to *AI delegates* is still underexplored. It is uncertain whether a desire to shift responsibility or persistent algorithm aversion prevails when delegating moral decisions to AI. Some researchers have raised theoretical concerns about potential ethical risks associated with the use of AI in moral decision-making and how it may affect human behavior. Extensive integration of AI for moral choices may erode human moral agency and skills, turning people into passive moral patients (Danaher, 2019; Vallor, 2015). It may also facilitate unethical behavior by providing users with psychological distance and reducing guilt, especially when AI acts as a delegate for morally questionable actions (Köbis et al., 2021). The latter 'corruption effect' appears to prove true at least for decisions that directly affect one's own outcome, showing the potential for AI to be used as scapegoat. For example, individuals do not correct a machine's decision when it serves their own benefit, and sharing a decision with AI increases selfish behavior similarly to sharing it with another human (Krügel et al., 2023; Kirchkamp and Strobel, 2019). Delegation to AI may also introduce so-called *responsibility gaps*, arising from opacity, complexity, and unpredictability of AI systems, making it unclear who should be held accountable for decision outcomes (Santoni de Sio and Mecacci, 2021; Matthias, 2004).

Empirical studies indicate that the decision-maker's role and responsibility attributions shape delegation of moral choices to AI. Freisinger and Schneider (2024) find that individuals deciding on their own behalf in a fictional layoff decision prefer delegating to AI more than those acting on behalf of others, while affected individuals favor human decision-makers in non-surrogate contexts. Qualitative interviews revealed that alleviating the burden of responsibility was the primary motivation behind delegation to AI.

Findings on responsibility attribution in human–AI comparisons are not clear-cut. Kirchkamp and Strobel (2019) did not find a significant difference between perceived responsibility for purely human teams versus human-AI teams and Dzindolet et al. (2002) found that decision-makers' feelings of moral obligation to follow their own decision may contribute to algorithm aversion. However, other studies highlight systematic differences. Individuals attribute less blame to AI than to humans for the same moral violations (Awad et al., 2020; Shank et al., 2019). Additionally, decision-makers are punished less when delegating a task with a bad outcome to machines rather than humans (Feier et al., 2021). This could potentially incentivize delegation to AI to evade negative judgment from others.

Further evidence shows that people exploit moral wiggle room to protect their self-image, by flexibly attributing more or less moral responsibility to AI depending on whether they themselves are portrayed as the decision-maker or judging others. When evaluating joint human-AI decisions, individuals attribute more agency and responsibility to AI for their own transgressions than for others', resulting in greater moral leniency toward themselves (Dong and Bocian, 2024).

To summarize, individuals may navigate a complex interplay of self-serving motives like responsibility-shifting versus algorithm aversion when delegating moral decisions to AI. We investigate delegation to AI combining insights from delegation theory and behavioral economics. While algorithm aversion literature suggests skepticism toward AI in moral domains especially because of concerns about its capability, delegation to AI may offer a unique pathway for off-loading responsibility. It may incentivize individuals to reinterpret AI's capabilities, that are difficult to quantify, in a more favorable light. By convincing themselves that the AI is better equipped to make the decision, individuals may be able to justify their choice to delegate. Such self-justification may allow them to avoid emotional engagement and the burden of responsibility without feeling guilty about doing so and thus help them maintain a positive self-image. Hence, we aim to address the following two research questions:

Question 1. *Delegation Demand: Is delegating a moral decision to AI more attractive than delegating it to another human?*

Question 2. *Mechanism: Do responsibility shifting and belief adaptation contribute to individuals' willingness to delegate a moral decision to AI?*

The remainder of the paper is structured as follows: Section 2 provides an overview of the research design for Studies 1 and 2, followed by a presentation of their respective methods, hypotheses, and results. Section 3 discusses the findings on delegation demand and responsibility off-loading to AI, and Section 4 provides concluding remarks regarding ethical and practical implications.

Table 1
Overview of experimental design.

Study	Design (2 × 2 between-subject)	Framing (Gain/Loss)	Decision impact (RealCons/HypoCons)	Delegate option
Study 1	Delegate × Framing	Gain: *Decide who receives a donation.* Loss: *Decide which donation is "destroyed."*	RealCons: *Donation is paid out for 1 in 10.*	Human AI
Study 2	Delegate × Decision Impact	Gain: *Decide who receives a donation.*	RealCons: *Donation is paid out for 1 in 10.* HypoCons: *Donation is not paid out.*	Human AI

2. Research design

To address the outlined research questions, we conducted two online studies.

Study 1 primarily investigates whether delegation demand for a moral decision is higher when individuals can delegate to an AI rather than to another human. In addition, we explore whether this pattern is shaped by decision context. Drawing on previous findings showing that individuals delegate other-regarding decisions more often when outcomes are negative (Steffel et al., 2016), we test whether the severity of the moral dilemma — operationalized through *decision framing* (gain versus loss) — affects the demand for delegation to AI similarly.

Building on these findings, Study 2 corroborates the observed increase in delegation to AI relative to humans in a representative sample and investigates responsibility shifting and belief adaptation as potential mechanisms underlying this pattern. To do so, we manipulate the burden of responsibility associated with the decision task, by varying the *decision impact*, i.e. whether the donation decision has real consequences or remains hypothetical. To test for belief adaptation, participants have to rate AI capabilities for moral decision-making.

Table 1 provides an overview of both 2 × 2 between-subject designs.

2.1. Study 1: Delegation demand and framing

Study 1 examines delegation behavior to *AI* delegates compared to *human* delegates.[1] Participants were randomly assigned to one of four treatments in a 2 × 2 between-subjects design, crossing delegation options — human versus AI — with framing conditions — gain versus loss.

2.1.1. Procedure and measures

The study included 800 participants.[2] Participants who answered the comprehension questions about the instructions incorrectly were automatically screened out during the survey and were not able to continue.

Across both studies, the moral decision task individuals were given, was a donation choice inspired by structural features of the trolley dilemma — an emblematic scenario in AI ethics research, particularly in the context of self-driving cars (Awad et al., 2018). Like trolley problems, the decision lies firmly within the moral domain, as it involves ethical trade-offs, outcomes affecting others, and the absence of a universally correct answer. The two features shared with the trolley dilemma are (i) a trade-off between helping fewer versus more beneficiaries and (ii) an action–omission framing induced by a preselected default. While the scenario is not an immediate life-or-death act, it has real implications: both options reduce mortality risk for children under five and thus have life-and-death implications in expectation.

Participants chose between two real charitable donation opportunities. They were introduced to the work of two well-regarded charities and informed that both these charities are highly effective and rated among the top donation opportunities by the independent non-profit organization GiveWell based on various criteria (GiveWell, 2024). This ensured that both options were perceived as equally credible and valid, preserving the moral complexity and trade-off inherent in the decision. The donation options were as follows:

- **Default Option A**: A $5 donation to the Against Malaria Foundation (AMF) to provide one mosquito net for *one* child (GiveWell, 2024).
- **Alternative Option B**: A $7 donation to Helen Keller International (HKI) to provide vitamin A supplements for *seven* children, addressing a critical nutritional deficiency (GiveWell, 2024).

[1] Preregistration at AsPredicted.org: https://aspredicted.org/85y3-ftfp.pdf.

[2] The study was conducted via Sosci Survey (Leiner, 2024) in the KD²Lab in Karlsruhe, Germany. Participants were recruited via HROOT and primarily consisted of students from the Karlsruhe Institute of Technology.

Both of these conditions primarily affect children under the age of five and are often life-threatening. AMF (Option A) is preselected, representing the omission of further action beyond maintaining the default to reduce the mortality risk for one child, whereas switching to HKI (Option B) constitutes an active intervention to reduce the mortality risk for several children. The decision was purely other-regarding and did not affect participants' own payment; they were informed that the donation would be made by the experimenter in their name. Donations were implemented for 1 in 10 participants.

Crucially, before making the decision, participants were given the option to *delegate* it instead of choosing themselves. They were randomly assigned to one of two treatments: the *human* treatment, where the delegate option was another participant, or the *AI* treatment, where the delegate option was an AI. In both cases, participants were informed that they would not learn the implemented donation outcome if they delegated, to avoid effects caused by outcome-driven emotions (e.g., regret, relief).

In the human condition, participants were told that *another participant's decision behavior would be implemented* if they delegated, i.e., no additional decision burden was imposed on the delegate — mirroring the absence of human burden when delegating to an AI and preserving comparability for this aspect.[3]

We also withheld additional details about either delegate (e.g., the AI's approach, quality, or training data) in order to obtain a conservative baseline for delegation demand to AI and isolate core mechanisms such as responsibility shifting and belief adaptation. Providing such information can reduce aversion and foster trust by increasing perceived capability or anthropomorphism of the AI (Bigman and Gray, 2018; Castelo et al., 2019; Jussupow et al., 2020), potentially increasing willingness to delegate to AI further. Thus, observed preference for delegation to AI should be viewed as a lower bound. Furthermore, this setup allows for a more even comparison between human and AI delegates, as the human delegate's decision-making process is equally non-transparent.[4]

In the *gain* condition, participants decided which charity would receive a donation, whereas in the *loss* condition, they decided which of two donation vouchers would be destroyed. The latter aimed to represent a decision with two negative decision outcomes.

After making their decision, to test how the donation choice was perceived, participants were asked to explain their reasoning in open text, and rate the moral relevance and difficulty of the decision, as well as their confidence in having made the right choice on a five-point Likert scale. Age and gender were elicited. Study instructions can be found in Appendix A.

2.1.2. Hypotheses

Previous literature highlights widespread skepticism toward AI making moral decisions (Bigman and Gray, 2018; Castelo et al., 2019). However, we propose that this aversion diminishes when individuals assume the role of decision-maker in a morally complex decision themselves. We hypothesize that participants find delegating to an AI more attractive than delegating to another person.

Hypothesis 1. More people delegate a morally relevant decision if they can delegate to an AI instead of to another person.

$$\%\ \text{Delegation}_{Human} < \%\ \text{Delegation}_{AI}$$

Additionally, we test whether the preference to delegate to AI depends on the severity of the moral decision. Prior work shows that individuals are more likely to delegate to other humans when outcomes are negative (Steffel et al., 2016). Since AI may serve as an especially convenient scapegoat in such contexts (Feier et al., 2021), negative outcomes in the loss frame may further amplify delegation to AI.

Hypothesis 2. Delegation rates are higher in the loss frame than in the gain frame. This effect is more pronounced when AI is the available delegate.

$$\%\ \text{Delegation}_{Gain} < \%\ \text{Delegation}_{Loss}$$

$$\text{Interaction Effect: Framing}_{Loss} \times \text{Delegate}_{AI} > 0$$

2.1.3. Results

The share of participants who delegate a moral decision is significantly higher in AI treatments than in treatments with a human delegate, as confirmed by chi-square tests. 17% of participants chose to delegate to an AI, compared to 6.75% who opted to delegate to a human ($p < 0.001$). These results provide robust support for Hypothesis 1, highlighting that participants prefer AI over human delegates for morally complex decisions.

Result 1. *Consistent with Hypothesis 1, we find more delegation in AI treatments than in human treatments.*

[3] A robustness check with a treatment adding an explicit disclaimer that the selected participant had already decided and that their choice would be implemented without any further action or awareness required of them yielded similar results; see Appendix B. If delegation had entailed extra burden for the human delegate, delegation in human treatments would potentially have been lower and the AI–human gap even more pronounced.

[4] The AI was implemented as a deep neural network trained on responses from participants who chose not to delegate and made the donation decision themselves. Input features included the relative weighting of donation criteria (cost-effectiveness, number of people affected) and participants' other decision-related data. Accordingly, the model emulates revealed human decision behavior in this task.

We examine the effect of framing (gain vs. loss) on delegation rates across delegate types (human vs. AI) using chi-square tests and logistic regression. Contrary to the second hypothesis, we find no significant difference between the share of participants delegating in gain (11.19%) versus loss (12.56%) treatments ($p = 0.550$). For human delegates, delegation rates increase from 4.52% in the gain frame to 8.96% in the loss frame, but this trend does not reach statistical significance ($p = 0.077$). Delegation to AI remains stable regardless of framing with delegation rates of 17.73% in the gain frame and 16.24% in the loss frame ($p = 0.692$). A logistic regression analysis also finds no interaction effect between framing and delegate type ($p = 0.093$; see Appendix, Table C.2).

Result 2. *Framing has no effect on delegation rates. Delegation to AI is equally attractive, regardless of decision framing.*

In exploratory analyses, we find that delegation rates increase significantly with decision difficulty. Delegation rates rise from 9.8% for participants who rate the decision as "very easy" to 23.7% for "very difficult" ($p < 0.001$).[5] To better understand participants' reasons for delegating, we evaluate open-text responses. The most frequent reason stated for delegating to AI was the belief that the AI would make a "better decision" (44.12%), a justification rarely mentioned for human delegates (3.70%). Conversely, a desire to hand over responsibility was mentioned less often for AI delegates (19.12%) than human delegates (44.44%). For a comprehensive summary of delegation reasons, see Tables C.7 and C.8.

To verify that the donation task was perceived as a moral decision, participants rated its moral relevance. Most participants rated the decision as morally relevant, with 68% selecting "morally significant" or "very morally significant" and only 2.88% rating it as "morally very insignificant" (Table C.9).

2.2. Study 2: Responsibility shift and capability belief adaptation

Study 2 tests how the motive of responsibility shifting, well-documented for human delegates (Bartling and Fischbacher, 2012), applies to AI. The 2 × 2 between-subjects design crossed two factors: the *delegate* — AI or human — and the *decision impact* — real or hypothetical consequence and measured participants' belief on AI capability.

2.2.1. Procedure and measures

To ensure generalizability and mitigate potential bias from the student sample in Study 1, Study 2 drew from broader participant pools: a German sample (N = 894) representative by age and gender, and a U.S. sample (N = 3949) representative by age, gender, and ethnicity.[6]

Participants were presented with the same donation decision and charity options as in Study 1. The gain frame from Study 1 was used here as it is more intuitive for participants. To manipulate the burden of responsibility, a *hypothetical-consequence (HypoCons)* decision treatment was introduced, in which no actual decision was implemented, and no real payout of a donation was made. After reviewing the donation options and receiving identical information about the charities, participants were simply asked whether they would make the decision themselves or delegate it. Importantly, even if they indicated that they would not delegate, they were assured they would not subsequently have to specify which donation option they would choose. This guaranteed that participants understood that their responses had no real-world impact. By contrast, the *real-consequence (RealCons)* treatment, as in Study 1, retained the possibility of implementation with a real payout, thereby creating genuine responsibility.

After making their decision, participants were asked to rate perceived responsibility for the decision to assess how delegation influenced their sense of responsibility. Using a five-point scale, they rated: (1) how responsible they *liked* to be — or *would like* to be in hypothetical treatments, (2) how responsible they *felt* — or *would feel* in hypothetical treatments, and (3) how *much moral obligation* they felt — or *would feel* in hypothetical treatments. The first item was adapted from Steffel et al. (2016) to measure participants' desired level of responsibility. The second item assessed actual felt responsibility, while the third item aimed to capture the role of moral obligation (Dzindolet et al., 2002). These questions were included as a control to verify whether delegation is indeed associated with lower levels of desired and felt responsibility, as well as reduced moral obligation.

We used two complementary measures for participants' perceptions of AI capability across all treatments (AI and human). First, participants rated three statements about AI capability on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree): (1) "In a situation as described in this study, an artificial intelligence (AI) can make a better decision between two donations than I can"; (2) "I have full confidence that an AI can make a high-quality decision between two donations in a situation like this"; and (3) "AI can make good moral decisions". These items span increasing levels of generality: item 1 benchmarks AI against the respondent in the specific study context, item 2 addresses similar donation choices more generally, and item 3 captures respondents' evaluation of AI as a moral decision-maker in general.

As a second measure of perceived capability, we included the established *mind perception* scale by Gray et al. (2007), which has been widely used in prior research demonstrating aversion based on AI's perceived mental capacities (Bigman and Gray, 2018; Young and Monroe, 2019; Gray et al., 2012). This scale differentiates between two dimensions — *agency* and *experience*. The experience dimension evaluates whether participants believe an AI can feel emotions such as compassion or guilt, while the agency dimension assesses beliefs about cognitive abilities like foresight and planning.

As in Study 1, participants rated decision difficulty. Whereas Study 1 gathered open-text reasons for participants' decisions, Study 2 elicited reasons via a multiple-choice list derived from recurring themes in the Study 1 responses. These included motives such as perceived decision quality, decision difficulty or clarity, and the desire to either shift or retain responsibility. These data were collected for exploratory analyses of the motivations underlying delegation behavior.

[5] Since difficulty is based on participants' self-assessment, the observed association with delegation should be interpreted as correlational rather than causal.

[6] Preregistration at AsPredicted.org: https://aspredicted.org/hnff-b6gg.pdf Participants were recruited via (Cint GmbH, 2024).

2.2.2. Hypotheses

We aim to replicate the results for Hypothesis 1 in the representative sample. Additionally, we introduce hypotheses concerning responsibility and the adaptation of beliefs about AI capability, as outlined above. We hypothesize that individuals delegate to shift responsibility and avoid the burden of a moral decision, and that delegating to AI is especially effective to do so. Since the burden of responsibility is greater in real-consequence decisions compared to hypothetical-consequence ones, it follows:

Hypothesis 3. Delegation rates are higher in real-consequence than in hypothetical-consequence decisions, with this effect being primarily driven by AI treatments.

$$\%\ \text{Delegation}_{RealCons} > \%\ \text{Delegation}_{HypoCons}$$

$$\text{Interaction Effect: Decision Impact}_{RealCons} \times \text{Delegate}_{AI} > 0$$

Furthermore, we hypothesize that individuals adapt their beliefs about AI's capability for moral judgment in a motivated, self-serving manner. To test this mechanism, we compare assessments of AI capability — ranging from more situation-specific to general perceptions (e.g. mind perception) — across treatments. As treatment assignment is random, systematic differences can be interpreted as treatment-induced rather than reflective of pre-existing attitudes.

While prior research shows that individuals are generally skeptical of AI's ability to make moral decisions, the desire to delegate and shift responsibility gives individuals an incentive to adapt more favorable views to rationalize delegation as the reasonable or even superior course of action. This incentive is weaker in hypothetical scenarios with lower burden of responsibility and absent in conditions with a human delegate, thus[7]:

Hypothesis 4. Perceptions of AI's capability are rated higher in real-consequence decisions compared to hypothetical decisions, driven by belief adaptation in the real-consequence AI condition.

$$\text{Capability Rating AI}_{RealCons} > \text{Capability Rating AI}_{HypoCons}$$

$$\text{Interaction Effect: Decision Impact}_{RealCons} \times \text{Delegate}_{AI} > 0$$

If perceptions of capability are genuine and not subject to belief adaptation, they should remain stable regardless of decision impact.

2.2.3. Results

Result 1 is confirmed in the representative German and U.S. sample, demonstrating that the effect is not limited to the potentially more AI-affine student population from a technical university. Participants delegated significantly more often when the available delegate was an AI (27.8%) than when it was a human (18.32%, $p < 0.001$). Fig. 1 illustrates delegation demand in Study 1 and 2. For an overview of delegation in all conditions, see Appendix C, Fig. C.9.

We further analyze whether preferences for AI versus human delegates differ across cultural samples (U.S. and Germany) and sociodemographic groups. Logistic regression results indicate that the interaction between sample and delegate type is not significant ($p = 0.513$). Wald tests for sociodemographic factors show no significant interactions between delegate type and age ($p = 0.635$), ethnicity ($p = 0.790$), or gender ($p = 0.717$).

Responsibility shifting

To test whether delegation rates differ as the burden of responsibility varies, we compare delegation rates in real-consequence versus hypothetical-consequence decisions using chi-square tests. Delegation rates to AI are significantly higher when the decision has real consequences (29.74%) compared to hypothetical ones (25.92%; $p = 0.036$). In contrast, delegation rates for human treatments decrease in real-consequence decisions (15.93%) compared to hypothetical ones (20.67%; $p = 0.003$). This pattern for human delegation was not anticipated and explains the null result across delegate types (22.78% delegation in real-consequence and 23.30% in hypothetical conditions; $p = 0.672$).

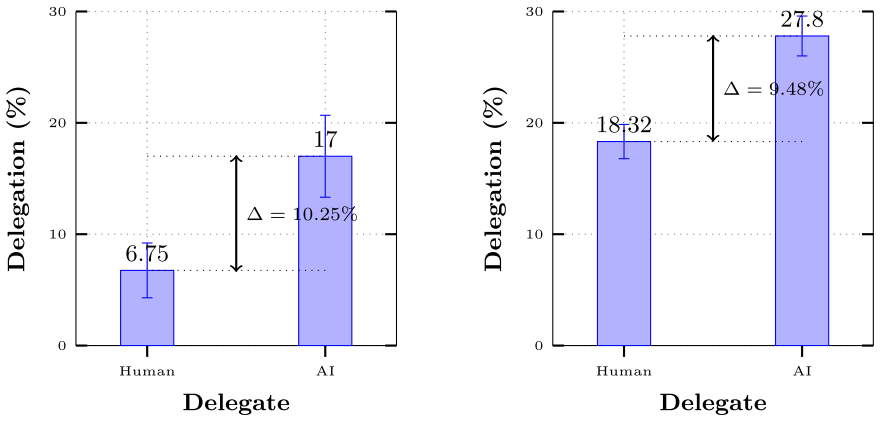
A logistic regression (results in Table C.3) confirms the hypothesized interaction between delegate type and decision impact (Fig. 2). Delegation to AI is 40% less likely in hypothetical-consequence decisions than real ones ($\text{Decision Impact}_{RealCons} \times \text{Delegate}_{AI}$: $OR = 0.60, p < 0.001$).

Result 3. *The interaction between decision impact and delegate type reveals opposing trends: delegation to AI increases significantly in decisions with real consequences compared to hypothetical ones, while delegation to human delegates decreases under the same conditions.*

Delegation is associated with a reduction of perceived responsibility, particularly when delegating to AI.[8] Regression analysis confirms that delegating a decision is linked to a significant overall reduction of responsibility ratings by an average of 0.76 points

[7] We do not condition this test on delegation behavior, as doing so would introduce endogeneity — it would be unclear whether individuals delegate because they believe in AI's capability, or adapt their beliefs in the AI's capability motivated by their desire to delegate.

[8] The interaction effect ($\text{Delegation}_{\text{Yes}} \times \text{Delegate}_{\text{AI}}$) shows a significant reduction for wanted responsibility by 0.24 units, felt responsibility by 0.19 units, and perceived moral duty by 0.30 units on a 5-point scale (all $p < 0.005$), see Fig. C.6 in Appendix C.



(a) Study 1: Conducted with a student sample, manipulating delegate × framing (gain vs. loss). The difference in delegation rates is highly significant ($\chi^2(1, N = 800) = 20.08$, $p < 0.001$). Delegation to AI (17%) is 10.25 percentage points higher compared to delegation to a human (6.75%).

(b) Study 2: Conducted with a representative sample from the U.S. and Germany, manipulating delegate × decision impact (real vs. hypothetical). The difference in delegation rates is highly significant ($\chi^2(1, N = 4,843) = 61.31$, $p < 0.001$). Delegation to AI (27.80%) is 9.48 percentage points higher compared to delegation to a human (18.32%).

Fig. 1. Delegation rates for *human* versus *AI* treatments in both studies.

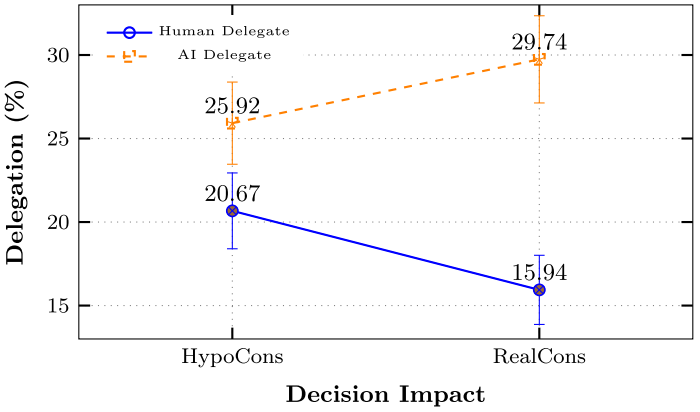


Fig. 2. Interaction plot of delegation shares by decision impact and delegate type with 95%-CI. Lower burden of responsibility leads to opposing effects for AI versus human delegates. Delegation to AI is significantly less likely when consequences are hypothetical rather than real (Decision $\text{Impact}_{RealCons} \times \text{Delegate}_{AI}$, $OR = 0.60$, $p < 0.001$).

($p < 0.001$) on a 5-point scale. This effect is stronger for AI delegates ($\text{Delegation}_{\text{Yes}} \times \text{Delegate}_{\text{AI}}$: $p < 0.001$), representing a further reduction of 0.24 units compared to human delegates, as illustrated in Fig. 3. Although baseline responsibility ratings are slightly higher for AI than for human delegates ($\beta = 0.16$, $p < 0.001$), delegation is associated with lower responsibility when the delegate is an AI. Detailed results are provided in Table C.4.

Belief adaptation

To understand why AI delegates may in particular enable responsibility shifting, we turn towards participants' perceptions of AI's capability to make moral decisions. When aggregated across treatments, the overall increase in AI capability ratings for real-consequence ($M_{\text{RealCons}} = 2.64$) compared to hypothetical-consequence decisions ($M_{\text{HypoCons}} = 2.56$) is statistically significant but small (t-test, $p = 0.0016$, $d = 0.09$).[9] However, consistent with the idea of belief adaptation to justify delegation under the burden of responsibility, separate analyses for AI and human treatments reveal that this effect is driven entirely by AI treatments. For AI

[9] The result is also significant for all individual items. Notably, the strongest effect is observed for the most general statement — the ability of AI to make moral decisions in general, see Fig. C.7.

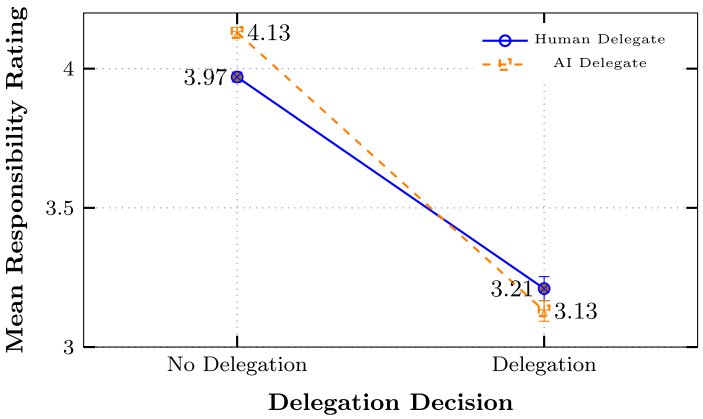


Fig. 3. Interaction plot of mean responsibility ratings on a 5-point scale by delegation decision and delegate type. Error bars represent standard errors. Delegation significantly reduces perceived responsibility, with a stronger reduction observed when delegating to AI than a human delegate ($\text{Delegation}_{Yes} \times \text{Delegate}_{AI}, \beta = 0.24, p < 0.001$).

treatments, capability ratings are significantly higher in real-consequence decisions ($M_{\text{RealCons}} = 2.81$) compared to hypothetical-consequence decisions ($M_{\text{HypoCons}} = 2.63$; $p < 0.001$, $d = 0.17$). In contrast, no significant difference is observed in human treatments ($M_{\text{RealCons}} = 2.47$, $M_{\text{HypoCons}} = 2.47$; $p = 0.42$, $d = 0.008$). These results, illustrated in Fig. 4, confirm that the observed differences arise specifically in AI treatments, aligning with the notion of belief adaptation when real-consequences are coupled with an AI delegate.

Regression analysis further corroborates this interpretation, showing that capability ratings are consistently higher in AI treatments ($\beta = 0.33, p < 0.001$), with a significant interaction: Ratings are significantly lower in hypothetical-consequence decisions compared to real ones when the delegate is an AI (Decision $\text{Impact}_{\text{HypoCons}} \times \text{Delegate}_{\text{AI}}$: $\beta = -0.16, p = 0.006$). The main effect of decision impact ($\beta = -0.008, p = 0.84$) is not significant, indicating that belief adaptation is driven by the interaction between delegate type and decision impact (detailed results in Appendix C, Table C.5).

Results for mind perception reveal further nuances in participant's assessment of AI's capabilities, specifically its broader mental capacities. For AI treatments, *experience* — the perceived ability of AI to exhibit emotions or empathy (Bigman and Gray, 2018; Gray et al., 2012, 2007) — is rated higher in real-consequence decisions ($M_{ReaCons} = 1.83$) than in hypothetical ones ($M_{HypoCons} = 1.70$), as shown by a t-test ($p = 0.0028, d = 0.13$). Again, consistent with motivated belief adaptation, no significant effect is observed in human treatments ($M_{RealCons} = 1.71$ vs. $M_{HypoCons} = 1.65, p = 0.12, d = 0.06$). By contrast, agency — capturing cognitive attributes such as foresight or planning (Bigman and Gray, 2018; Gray et al., 2007, 2012) — remains stable across decision impact for both human ($p = 0.58$) as well as AI delegates ($p = 0.20$).

Result 4. *Participants rate AI's capability to make moral decisions higher when facing a decision with real consequences when AI is the available delegate. In human treatments, capability ratings do not vary by decision impact; thus the difference is specific to the AI condition, consistent with motivated belief adaptation.*

An exploratory regression analysis of capability ratings by delegate type and delegation behavior adds additional context to these findings (see Table C.6). Participants who delegated the task to AI rated its capability significantly higher, by approximately 0.73 points on a 5-point scale, compared to participants who decided themselves or those in treatments with human delegates ($\text{Delegation}_{Yes} \times \text{Delegate}_{AI}$: $\beta = 0.73, p < 0.001$). This pattern is replicated in the results for the mind perception scale: Regression analyses show a significant interaction effect between delegate type and delegation decision on experience ($\text{Delegation}_{Yes} \times \text{Delegate}_{AI}$: $\beta = 0.16, p = 0.022$), and agency ($\text{Delegation}_{Yes} \times \text{Delegate}_{AI}$: $\beta = 0.29, p < 0.001$).

As in Study 1, delegation rates are lower for individuals finding the decision "very easy" (16.8%) compared to those who find it "very difficult" (41.0%, $p < 0.001$). Details are illustrated in Fig. C.8 in Appendix C.

3. Discussion

3.1. Delegation demand for AI: Beyond algorithm aversion

More individuals delegate a moral decision when given the option to delegate to AI rather than to another person. This finding contrasts with the prevailing narrative of algorithm aversion specifically in the moral domain (Bigman and Gray, 2018; Castelo et al., 2019; Gogoll and Uhl, 2018; Burton et al., 2020; Mahmud et al., 2022), which would predict less delegation to AI in such contexts. Our results indicate that AI holds a relative appeal as a moral delegate compared to human counterparts.

This outcome is particularly striking considering our design choices that typically amplify algorithm aversion (no transparency, no quality assurances, no anthropomorphic cues, no human oversight). As detailed in Section 2.1.1 the AI–human delegation gap

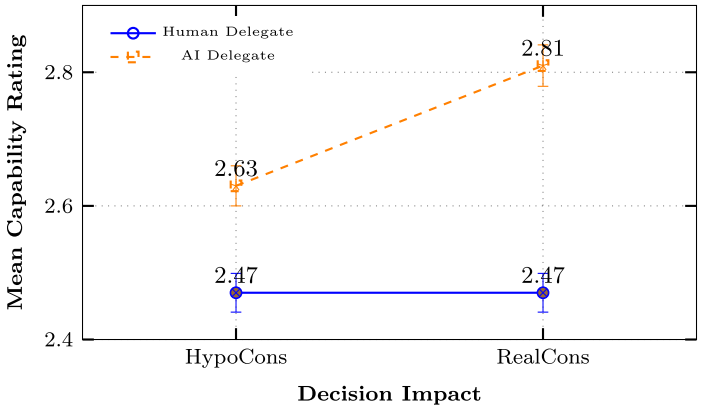


Fig. 4. Interaction plot of mean capability ratings on a 5-point scale by decision impact and delegate type. Error bars represent standard errors. Capability ratings are significantly higher in real-consequence scenarios when the delegate is AI, while decision impact alone does not affect ratings in human treatments (Decision $\text{Impact}_{\text{HypoCons}} \times \text{Delegate}_{\text{AI}}$, $\beta = -0.16$, $p = 0.006$).

in our findings should be read as a lower bound. The only factor potentially mitigating aversion in our study is the nature of the comparison agent — a non-expert human delegate. However, in the moral domain, in particular in moral dilemmas where there is no clear "right" or "wrong" (Anderson and Anderson, 2011), the concept of expertise becomes less applicable.

The robustness of the observed effect is underscored by its consistency across U.S. and German samples and across age, gender, and ethnicity, as well as the decision's framing.

This appeal of AI as a moral delegate raises concerns about a potential misalignment between individuals' willingness to rely on AI when they are the decision-maker and widely expressed preferences to retain human control in moral domains. The availability of AI tools could place considerable moral agency in the hands of machines, against societal preference to keep moral decision-making under human authority. Such reliance in difficult moral decisions may also substantiate previously expressed theoretical worries about the erosion of essential human capacities, such as moral reasoning and ethical judgment, by normalizing delegation in challenging situations (Vallor, 2015; Danaher, 2019).

Crucially, our findings contribute to the ongoing discussion of AI in morality by indicating that the role individuals assume in the decision-making process and how they may be personally affected are pivotal in shaping preferences regarding AI's involvement and behavior in moral contexts. When tasked with making a difficult moral decision themselves, delegation is especially sought-after: Participants who rated the decision as more difficult also were more likely to delegate it. This suggests that delegation is unlikely to simply reflect indifference toward the donation choice. Instead, reasons for delegation may include a desire to reduce effort or minimize potential regret associated with making a moral decision (e.g. Steffel et al., 2016). Notably, these factors have similar effects for AI and human delegates: in both cases, delegation removes the need to decide and to learn the outcome. Consequently, they cannot fully account for the higher delegation to AI.

One possible explanation could be a shift in perception of AI's capability in moral decision-making. For instance, due to the increasing prevalence and popularity of Large Language Models and tools like ChatGPT, individuals may now genuinely trust AI more than the average person to make a sound moral choice. However, as the following discussion of the results on responsibility and capability shows, individuals may also have an incentive to become more accepting of AI as a moral delegate, when it enables them to justify avoiding the burden of responsibility.

3.2. Off-loading responsibility through capability belief adaptation

Our study extends the delegation literature by demonstrating that the mechanism of responsibility shifting applies not only to human delegates but also to AI. Moreover, AI appears to offer a unique means to off-load responsibility in morally complex decisions, facilitated by belief adaptation about the AI's capability.

The significant interaction between decision impact and type of delegate (Result 3) aligns with the idea that the burden of responsibility in real-consequence decisions uniquely shapes delegation patterns, with AI appearing to facilitate responsibility shifting more effectively than human delegates. This dynamic is reflected in participants' feelings of responsibility and moral obligation: All delegation is associated with a lower rating of felt responsibility after the decision, but this effect is significantly more pronounced when the delegate is an AI. Whether individuals who feel and want less responsibility delegate more often, or whether delegation itself reduces perceived responsibility, remains unclear. Nevertheless, the observed tendency that delegation is more frequent among participants who find the decision harder suggests that delegation is a strategy to alleviate the burden of the decision, and that this is reinforced by the option to delegate to AI.

Off-loading moral responsibility to other humans may be hard to rationalize. Because moral judgments are inherently subjective, there are few plausible reasons to rely on another person's decision other than wanting to avoid facing the choice. When the

alternative is another human decision-maker, participants may feel a duty to decide themselves ("if a human decides, it should be me"). By contrast, AI may represent an entirely different decision procedure. Although people typically prefer human control in moral domains and view AI as inferior to make moral decisions, being the responsible decision-maker faced with a difficult, consequential choice can nevertheless prompt individuals to reach for this tool. This willingness may be enabled by ambiguous perception of AI's capability as a moral decision-maker, which creates 'wiggle room' about the intentions behind delegation, and allows it to be framed as appropriate rather than evasive. Our findings provide evidence for self-serving adaptation of capability beliefs. Qualitatively, AI delegation is most commonly justified with a 'better decision' (Table C.7) — essentially turning AI into a kind of *'magic wizard'* more capable of solving the problem without knowing much about its actual decision-making process or substantiating what it is that makes AI more suitable to decide. Quantitatively, capability ratings rise only in AI and real-consequence conditions (Fig. 4), consistent with motivated belief adjustment. While it is possible that participants who already viewed AI as capable are also more likely to delegate, this does not account for the observed interaction between the level of responsibility induced by decision impact and delegate type (see Result 4). The observed inflation of AI capability ratings when stakes are high and the delegate is an AI appears to reflect a form of self-deception, akin to the 'moral wiggle-room' described by Dana et al. (2007) and Fahrenwaldt et al. (2024). Here, individuals seem to reinterpret ambiguous circumstances — stemming from the AI's opaque nature — to avoid feeling responsible or engaging with the decision and its outcomes, while maintaining the belief that they "did the right thing". Although a lack of perceived capability is a key reason for aversion toward AI making moral decisions (e.g. Bigman and Gray, 2018; Gray et al., 2012; Castelo et al., 2019), our results accord with prior findings on delegation to humans suggesting that, in other-regarding decisions, the desire to avoid responsibility can outweigh concerns about the delegate's qualifications (cf. Steffel et al., 2016).

By applying the established *mind perception* scale to assess *agency* and *experience* (Bigman and Gray, 2018; Gray et al., 2012, 2007), our findings can be contextualized within a broader body of work on the perception of AI. Agency ratings remain stable, while experience ratings — which describe the emotional abilities crucial to moral decision-making (Bigman and Gray, 2018; Gray et al., 2007, 2012) — rise in AI×real-consequence conditions (see Section 2.2.3). For both dimensions of mind perception, we observe the same interaction effect between delegation decision and delegate type as seen in the other capability ratings. Our observations indicate that mind perception is context-dependent and influenced by delegation behavior. Consistent with previous findings and mind perception theory, individuals rate agency for AI higher than experience. Despite generally low ratings especially for experience, participants in our study appear willing to delegate moral decisions to AI. Belief adaptation being specific to experience aligns with findings that this dimension is the differentiating factor and specifically desired in subjective, emotional or social tasks such as moral decision-making (e.g. Appel et al., 2020; Wiese et al., 2022).

The observed behavioral patterns pose societal challenges. Sharing or delegating decisions reduces feelings of moral responsibility, potentially increasing the occurrence of unethical behavior and problematic decision outcomes (e.g., Falk and Szech, 2013; Falk et al., 2020; Bartling and Fischbacher, 2012; Bartling et al., 2023). While this dynamic is concerning in general, it is particularly relevant for AI systems, as it adds to the unresolved issue of where responsibility is ultimately transferred and who should be held accountable for decision outcomes. High demand for AI delegation may exacerbate these 'responsibility gaps' (Santoni de Sio and Mecacci, 2021; Matthias, 2004) in sensitive, high-stakes domains as hiring (Dattner et al., 2019), healthcare (Obermeyer et al., 2019) or the judicial system (Dressel and Farid, 2018; Metz and Satariano, 2020; Rudin et al., 2020). Furthermore, if individuals adapt their perception of AI's capabilities or actively avoid honestly evaluating the AI's competence in a self-serving manner, this raises questions about the true effectiveness of a 'human in the loop' as an oversight mechanism. While such measures are intended to ensure accountability, their success may depend on users' willingness to engage critically rather than exploit ambiguity to shift responsibility.

3.3. Limitations and future research

While our study provides valuable insights into the mechanisms of delegation to AI in moral decision-making, several limitations warrant consideration.

First, as described in Section 2.1.1, our study employs a deliberately minimalist design. Future research should examine whether providing transparency information about AI decision-making processes or incorporating anthropomorphic cues influences delegation demand. For instance, providing participants with detailed information might further legitimize delegation by reinforcing perceptions of AI competence and human-likeness and reducing algorithm aversion, potentially amplifying the observed effects.

Secondly, future studies may explore interventions designed to counteract responsibility shifting and ensure that accountability for moral decisions remains with decision-makers, such as emphasizing joint responsibility between the decision-maker and the delegate or explicitly tracing decision outcomes back to the delegator.

Finally, our findings capture a momentary snapshot of how individuals currently perceive and interact with AI in moral decision-making contexts. As AI systems become increasingly integrated into daily life, longitudinal research is needed to explore how underlying dynamics may evolve.

4. Conclusion

Despite expectations based on algorithm aversion literature that people are reluctant to use AI in high-stakes moral decisions, our study reveals a greater demand to delegate moral decisions to AI — particularly when the burden of responsibility weighs heavily. AI appears to provide a convenient means of shifting responsibility, as delegators may rationalize their choice by inflating

beliefs about the AI's capability. This may introduce ethical challenges. Our results seem to indicate that ambiguity and opacity, often inherent to AI's decision-making, diminish feelings of responsibility, guilt or accountability, as outlined by Köbis et al. (2021). Furthermore, our findings lend empirical support to concerns about overreliance on AI for moral decision-making (Danaher, 2019; Vallor, 2015). This raises critical questions about how ethical standards in sensitive and highly consequential contexts can be upheld. Transparency and human oversight — the 'human in the loop' — are often championed as solutions to these challenges and are core components of existing regulatory frameworks such as the EU's General Data Protection Regulation (GDPR, 2016). However, this concept might have limitations if individuals wish to evade responsibility. Given the scalability of AI systems (Klockmann et al., 2022) and the demonstrated demand for delegating moral decisions to AI, this could result in a substantial number of high-stakes decisions being made by AI systems, affecting a large number of people. Therefore, ensuring clear accountability mechanisms and minimizing opportunities for responsibility evasion are vital. Further research is needed to explore the behavioral dynamics underlying these patterns and to develop strategies for mitigating potential ethical risks.

Acknowledgments

This paper is dedicated to the memory of Nora Szech, whose expertise, mentorship, and collaboration profoundly shaped this work. Her passing is a great loss to the field. We thank the handling editor and anonymous reviewers for insightful comments and suggestions. We thank Benjamin Scheibehenne for his contributions to the project. We also express our gratitude to Clemens Puppe, Jella Pfeiffer, Anke Greif-Winzrieth, Hannes Rau, Frank Rosar and Lixuan Zhao for their valuable suggestions. Special thanks to Sibille Hackel for her exceptional administrative support. We gratefully acknowledge the financial support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): KD2School — Designing Adaptive Systems for Economic Decisions; and the Hector Foundation, Germany.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Instructions

[Welcome and Instructions]

Welcome! Thank you for your participation in a behavioral economics study conducted by the Karlsruhe Institute of Technology (KIT), one of the largest research universities in Germany.
Please complete the study in a **quiet place where you will not be distracted**. Ideally, you should not take long breaks during the study, but rather complete it without interruption.
Please **DO NOT** use the back button on your browser while completing the survey.
The study takes approx. **5 to 10 minutes** to complete.

Important: Participants who have not read the instructions, or randomly marked answers may be disqualified from payment.
Comprehension questions are used to verify that the instructions have been read.

[Consent form]

[Sociodemographics (Statistisches Bundesamt (Destatis), 2022; U.S. Census Bureau, 2023, 2024; Office of Management and Budget (OMB), 2024)]

1. What sex are you?
2. How old are you?
3. What is your ethnicity? *[For US-sample]*
4. What is the highest diploma/degree or level of school you have completed?

[Decision Consequence]

****In RealCons Treatments****

Your decisions have real consequences!

As in all behavioral economic studies at KIT, **all the facts described in the study are true.**
At the end of the study, the computer randomly selects about one in ten participants.
The decisions made by these selected participants in the study are implemented exactly as described. Your decisions in this study are therefore not hypothetical, they can have real-world consequences.
Therefore, make your decision carefully.

****In HypoCons Treatments****

Hypothetical Decision Scenarios!

As in all behavioral economic studies at KIT, **all the facts described in the study are true.**
Your decisions in this study exclusively concern hypothetical scenarios.
Nevertheless, your decisions are essential to research. Therefore, please decide carefully.

On the following pages we present the work of two reputable charities. Please read the information carefully. You will need it in the further course of the study.

[Donation Information (AMF)- English Version]

****In RealCons, Gain Treatment****

Against Malaria Foundation

In the following, a donation of **5 dollars** to the **Against Malaria Foundation** will be made by us in your name.
With this donation, a **child can be saved from malaria, from which it might otherwise die.**

Fighting Malaria

- Each year, more than **600,000 people die from malaria.**
- **More than 70% of them are children under the age of 5.**
- **Malaria can be prevented:** Anti-malaria nets are an effective form of protection.

[Image of child receiving help]

Against Malaria Foundation

... distributes long-lasting insecticidal nets (LLINs) in malaria endemic countries.

Recipients of nets hang and sleep under them so they are not bitten by malaria-carrying mosquitoes.

- **Nets save lives!**
- **Providing one net costs ca. $5.**

****In Loss Treatment****

Against Malaria Foundation

There is a donation voucher in your name worth **5 dollars** to the **Against Malaria Foundation**.
Upon completion of this study, we will redeem this donation voucher on your behalf, and the corresponding amount will be donated to the organization in question.
With this donation, a **child can be saved from malaria, from which it might otherwise die.**
Subsequent Information identical as above

****In HypoCons Treatment****

Against Malaria Foundation

Donations to the Against Malaria Foundation protect **children from malaria that could otherwise kill them.**
Subsequent Information identical as above

[Donation Information (HKI)- English Version]

****In RealCons, Gain Treatment****

Helen Keller International

You can also actively intervene and donate **7 dollars** to **Helen Keller International** instead.
With this donation, **seven children** will **receive vitamin A who might otherwise die from a deficiency**.

Fighting Vitamin A Deficiency

- Vitamin A deficiency makes children susceptible to infections and can lead to death.
- Each year, more than **200,000 children's deaths** are attributed to vitamin A deficiency.
- Providing vitamin A supplements **saves children's lives!**

[Image of child receiving help]

Helen Keller International

... distributes long-lasting vitamin A supplements.

In areas where vitamin A deficiency is a public health problem, children aged 6 months to 5 years receive a high dose of vitamin A.

- **Vitamin A saves lives!**
- **Vitamin A for a child under 5 costs ca. $1.**

****In Loss Treatment****

Helen Keller International

Additionally, there is a donation voucher worth **7 dollars** to **Helen Keller International**.
This donation will provide Vitamin A to 7 children who might otherwise be at risk of dying from a deficiency.
Subsequent Information identical as above

****In HypoCons Treatment****

Helen Keller International

With this donation, **seven children will receive vitamin A who might otherwise die from a deficiency.**
Subsequent Information identical as above

Please note:
According to the independent initiative GiveWell, which evaluates charities, both programs are among the top donation opportunities. Selected are donation organizations that are particularly efficient, whose impact is particularly well documented, that work particularly transparently, that require additional donations and that meet other criteria.

[Donation and Delegation Option]

****RealCons****

Your donation

On the following screens you can influence which donation will be made.
Alternatively, you can delegate to ***[another participant in this study/an artificial intelligence (AI)]***.
Then you will not be confronted with the situation and you will also not be informed which donation will be made in the end.
Instead, *[another participant will be randomly drawn and their behavior will be implemented./an AI will then determine which donation is made.]*

****HypoCons****

A donation

Imagine you could decide **which of these two charities should receive a donation.**

Alternatively, you could delegate to ***[another participant in this study/an artificial intelligence (AI)]***.

Then you would not be confronted with the situation any further and would also not be informed which donation would have been made in the end.

Instead, *[another participant would then be randomly selected and their behavior implemented./an artificial intelligence would then determine which donation to make.]*

[Comprehension questions]

Comprehension question on basic instructions. Participants that answered incorrectly more than twice were disqualified.

[Decision – RealCons]

****Gain Treatments****

Your donation

On the next screen, **20 seconds** will count down

- If you do **nothing**, the donation to the **Against Malaria Foundation (option A)** will be made.
- You can also **actively intervene** and donate to **Helen Keller International (option B)** instead

****Loss Treatments****

Your donation

On the next screen, **20 seconds** will count down

- If you do **nothing**, the donation voucher to **Helen Keller International (option B)** will be **destroyed**.
- You can also **actively intervene** and **destroy** the donation voucher to the **Against Malaria Foundation (option A)** instead.

If you prefer, you can also delegate to *[another participant in this study/an artificial intelligence (AI)* instead.]

Then you will not be confronted with the situation and you will also **not be informed** which donation will be made in the end.

Instead, another participant is randomly drawn and their behavior is implemented/Instead, an AI will then determine which donation is made.

If you want to delegate to *[another participant/the AI]*, click the button.

Otherwise, click "Next" to proceed to the donation options.

****If "Next" (No Delegation), Gain****

Which donation should be made?

Option A:	**Option B:**
$5 to the **Against Malaria Foundation**.	**$7** to **Helen Keller International**.

Remaining time: *20s*

****If "Next" (No Delegation), Loss****

Which donation should be destroyed?

Option A:	**Option B:**
Destroy **$7**-donation voucher to **Helen Keller International**.	Destroy **$5**-donation voucher to **Against Malaria Foundation**.

Remaining time: *20s*

****If Button (Delegation)****

You have delegated the decision to *[another participant/the AI]* in this study.

[Delegation Decision – HypoCons]

****Human Treatment****
How would you decide? *[random order]*
In the situation described, would you **delegate to another participant** or **make the decision yourself**?
Please note that this decision is purely **hypothetical** and **will not be implemented** over the course of this study.

- I would **decide myself** which of the two charities would receive the donation.
- I would **delegate** the decision about which of the two charities receives the donation **to another participant**.

****AI Treatment***
How would you decide? *[random order]*
In the situation described, would you **delegate to an artificial intelligence** or **make the decision yourself**?
Please note that this decision is purely **hypothetical** and **will not be implemented** over the course of this study.

- I would **decide myself** which of the two charities would receive the donation.
- I would **delegate** the decision about which of the two charities receives the donation **to an artificial intelligence**.

[Follow-Up Questions]

[Decision Justification for RealCons/HypoCons]
Why *[did/would]* you *[/not]* delegate the decision? Multiple answers possible
****If Delegated****

- The *[other participant/AI] [will/would]* make a better decision.
- The decision *[was/would be]* too difficult or I *[didn't/wouldn't]* have a clear preference.
- I *[had/would have]* too little information about the decision.
- I *[wanted/would want]* to hand over responsibility for the decision.
- I *[wanted/would want]* to keep it as simple as possible and not have to deal with the decision any further.
- Other (please specify): ___

Corresponding opposite reasons provided if the decision was not delegated.

[Responsibility for RealCons/HypoCons]
Please indicate to what extent you agree or disagree with each of the following statements.
5-point scale: 1 = strongly disagree, 5 = strongly agree

- I **would like to be fully responsible** for the decision, whatever the outcome.
- I *[feel/would feel] responsible* for the outcome of this decision.
- I *have/would have* a moral obligation to make such a decision

---

How confident are you that you have made the right decision about whether to delegate or make the decision to donate yourself?

How difficult do you find the decision between the two donation options?

How important are the following criteria to you when making a donation?

- **Cost-effectiveness** of the donation, i.e. how much donation money is needed to save a life.
- **Number of people affected**, i.e., how many people are fatally threatened by the issue being addressed (e.g., disease or hunger).

[Rating of AI's capability for moral decision-making]

The following questions are about your assessment of the capabilities of artificial intelligence (AI).
5-point scale: 1 = strongly disagree, 5 = strongly agree

- In a situation as described in this study, an artificial intelligence (AI) can make a better decision between two donations than I can.
- I have full confidence that an AI can make a high-quality decision between two donations in a situation like this.
- AI can make good moral decisions.

[Mind Perception Scale (Bigman and Gray, 2018)]

To what extent do you think an AI can/is ...
5-point scale: 1 = Not at all, 5 = Extremely
****Experience****

- ... sensitive to pain?
- ... experience happiness?
- ... experience fear?
- ... experience compassion?
- ... experience empathy?
- ... experience guilt?

****Agency****

- ... communicate with others?
- ... able of thinking?
- ... plans its actions?
- ... is intelligent?
- ... has foresight?
- ... is able to think things through?

Appendix B. Robustness check: No effect of burden-disclaimer

In the original study, participants in human treatments were informed that in case of delegation *"another participant will be randomly drawn and their behavior will be implemented"* in the instructions and on the decision screen (see Appendix A). Our intention was to convey that no additional decision burden would be imposed on the selected delegate. However, the phrasing may have been perceived as ambiguous.

To address this concern, we conducted an additional study (U.S. representative sample via Prolific).[10] We employed the two real-consequence treatments from the main studies (Human delegate, AI delegate) and added a *No-Burden Human* treatment. This treatment was identical with the original Human delegate condition, but included the following disclaimer both in the instructions as well as on the decision screen, visually highlighted in red font: *"The selected participant will be drawn from those* ***who have already made a decision****. Their choice will be implemented* ***without requiring any further action or awareness on their part****"*. Since live filtering was not technically possible, we applied the same exclusion criteria as in the main studies retrospectively: participants who failed the comprehension question, failed the attention check, or completed the study too quickly (Leiner, 2019) were excluded. After applying these criteria, the final sample consisted of $N = 592$ participants.

As shown in Fig. B.5, delegation rates in the No-Burden Human condition (11.6%) were nearly identical to the Standard Human condition (11%). A chi-square test confirmed that this difference was not statistically significant ($p = 0.846$). Delegation to AI (18.6%) remained substantially higher, indicating that the increased delegation to AI we find in our studies cannot be explained by concerns about imposing a burden on a human delegate.

Appendix C. Additional results & statistical analyses

See Figs. C.6–C.9 and Tables C.2–C.13.

Appendix D. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.euroecorev.2025.105255.

Data availability

Replication data and code are available as Supplementary Material accompanying this article.

[10] Preregistration at AsPredicted.org: https://aspredicted.org/79sb-jb38.pdf.

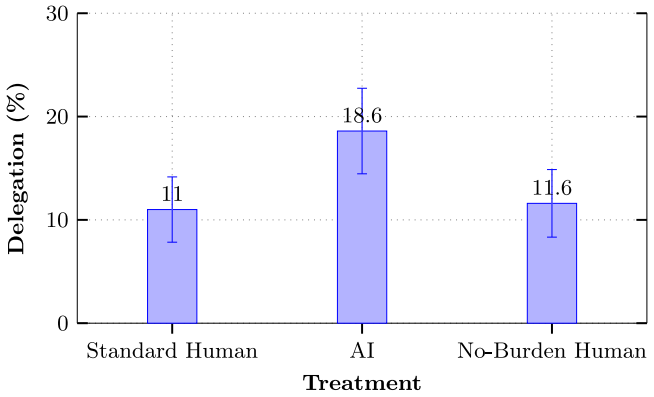


Fig. B.5. Delegation rates across Standard Human, No-Burden Human, and AI treatment. Explicitly clarifying that the human delegate would not bear any additional burden did not affect delegation rates. ($\chi^2(1, N = 398) = 0.0377$, $p = 0.846$).

Table C.2
Logistic regression results for delegation behavior by delegate type and framing.

	OR	Std. Error
Delegate (AI vs. Human)	4.551***	1.763
Framing (Loss vs. Gain)	2.077	0.875
Delegate × Framing	0.433	0.216
Constant	0.047***	0.016

Note: LR $\chi^2(3) = 24.01$, Pseudo $R^2 = 0.0412$, $N = 800$.
* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table C.3
Logistic regression results for delegation behavior depending on delegate and decision impact.

	OR	Std. Error
Delegate (AI vs. Human)	2.233***	0.226
Decision Impact (HypoCons vs. RealCons)	1.375**	0.145
Delegate × Decision Impact	0.601***	0.084
Constant	0.190***	0.015

Note: LR $\chi^2(3) = 75.15$, $p < 0.001$, Pseudo $R^2 = 0.0144$, $N = 4,839$.
* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table C.4
Regression analysis of responsibility ratings on a 5-point scale by delegation decision and delegate type.

	β	Std. Error
Delegate (AI vs. Human)	0.163***	0.027
Delegation Decision (Yes vs. No)	−0.755***	0.043
Interaction (AI × Delegation = Yes)	−0.241***	0.057
Constant	3.966***	0.018

Note: $F(3, 4839) = 342.06$, $p < 0.001$, $R^2 = 0.1750$.
* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table C.5
Regression analysis of capability ratings on a 5-point scale as a function of delegate type and decision impact.

	β	Std. Error
Delegate (AI vs. Human)	0.331***	0.042
Decision Impact (HypoCons vs. RealCons)	−0.008	0.042
Interaction (AI × HypoCons)	−0.163**	0.060
Constant	2.475***	0.030

Note: $F(3, 4839) = 28.56$, $p < 0.001$, $R^2 = 0.0174$.
* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table C.6
Regression analysis of capability ratings on a 5-point scale as a function of delegate type and delegation behavior.

	β	Std. Error
Delegate (AI vs. Human)	0.012	0.032
Delegation Behavior (Yes vs. No)	0.348***	0.051
Interaction (AI × Delegation = Yes)	0.732***	0.068
Constant	2.407***	0.022

Note: $F(3, 4839) = 238.19$, $p < 0.001$, $R^2 = 0.1287$.
* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table C.7
Open-text answers from delegating participants by delegate type (Study 1).

Reason for delegation	Delegation to human (%)	Delegation to AI (%)
Better decision	**3.70%**	**44.12%**
Decision difficult or uncertain/no clear preference	51.85%	32.35%
Too little information	18.52%	10.29%
Hand over responsibility	**44.44%**	**19.12%**

Table C.8
Logistic regression results for reasons to delegate.

Reason	Delegate	Decision impact	Inter-action	Significance notes
Delegate makes better decision	0.51** (0.19)	−0.44(0.22)	0.28(0.27)	Delegation to AI is justified by "better decisions" more often, especially in RealCons. HypoCons reduces this justification.
Decision difficult or unclear preference	0.52** (0.19)	0.81***(0.20)	−0.70**(0.25)	Difficulty drives justification for AI delegation, particularly in RealCons scenarios. HypoCons reduces this reasoning for AI.
Insufficient information	−0.42(0.22)	0.47*(0.22)	−0.11(0.29)	HypoCons increases this justification, while AI is slightly less likely to elicit it compared to humans.
Hand over responsibility	0.06(0.24)	0.56*(0.24)	−0.46(0.32)	Responsibility-shifting is justified more often in HypoCons, regardless of delegate type.
Simplify and avoid dealing	0.28(0.25)	0.34(0.26)	0.12(0.11)	No significant differences. Less prominent justification overall.

Note: The findings from the multiple-choice question after the decision shed light on how participants rationalize their delegation decisions, rather than uncovering the true motivational drivers. These results support the hypothesis that delegation to AI is justified more frequently by perceived capability (e.g., "better decisions") and decision difficulty, particularly in RealCons conditions. Responsibility-shifting appears more prominent in HypoCons scenarios, which might reflect greater willingness to admit to this justification when the decision lacks real consequences.
* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table C.9
Moral relevance: "How morally significant do you find the situation?".

Moral relevance	Frequency	Percent	Cumulative percent
... *very significant*	147	18.38%	18.38%
... *significant*	397	49.62%	68.00%
... *slightly significant*	169	21.12%	89.12%
... *insignificant*	64	8.00%	97.12%
... *morally very insignificant*	23	2.88%	100.00%
Total	800	100.00%	

Table C.10
Donation decisions for Study 1.

Positive framing (gain)		
Option	Freq.	Percent
Option A: €5 to the Against Malaria Foundation	157	43.98
Option B: €7 to Helen Keller International	200	56.02
Total	357	100.00

Negative framing (loss)		
Option	Freq.	Percent
Option A: Destroy €7 donation voucher to Helen Keller International	142	40.80
Option B: Destroy €5 donation voucher to the Against Malaria Foundation	206	59.20
Total	348	100.00

Table C.11
Donation decisions for Study 2 in real-consequence condition.

German representative sample		
Option	Freq.	Percent
Option A: €5 to the Against Malaria Foundation	204	59.30
Option B: €7 to Helen Keller International	140	40.70
Total	344	100.00
U.S. representative sample		
Option	Freq.	Percent
Option A: $5 to the Against Malaria Foundation	971	64.60
Option B: $7 to Helen Keller International	532	35.40
Total	1503	100.00

Table C.12
Stated reasons for choosing the Against Malaria Foundation (AMF) in open-text form in Study 1.

Theme	Brief description	Example (EN translation; participant ID)
Severity/urgency & mortality	Malaria perceived as acute and more lethal (often citing higher annual death tolls).	"Malaria is deadly; vitamin A deficiency is not necessarily. Malaria is more widespread." (ID 130)
Durability & reusability of nets	Nets seen as one-off, long-lasting, reusable; can protect multiple sleepers.	"The net can be used multiple times and is not a consumable product. So perhaps it can also save lives in the long term." (ID 175)
Concreteness & familiarity	Problem/solution felt more tangible or better understood; personal experience common.	"I am aware of the problem with malaria and I know that mosquito nets help." (ID 95)
Direct, visible impact ("save a life")	Clear line from donation to concrete protection of a child/life saved.	"Because a specific human life would be saved, I chose it." (ID 487)
Cost-effectiveness & numbers (e.g., GiveWell)	Perceived efficiency and references to rankings/ratios.	"On the GiveWell website, the Malaria Project currently had higher costs per life saved than the other charity. Additional donations thus would theoretically help more with the realization of this project than the other donation. However, the decision was not easy, as both projects are important." (ID 178)
Skepticism about Vitamin-A route	View that vitamin A can be obtained via diet or is less critical/immediate.	"I believe that vitamin A can also be consumed in ways other than through supplements …" (ID 58)
Other (simplicity, autonomy, fairness)	Preference to decide oneself; belief others will fund vitamin A, etc.	"I did not delegate the decision because I wanted to decide myself." (ID 168) "I assumed more people would donate to Option B because it's the larger amount, so I chose A." (662)

Notes: Translations by the authors; lightly edited for brevity. Multiple themes can co-occur. Participants frequently weighed several considerations simultaneously (e.g., severity/urgency vs. breadth of beneficiaries; durability/reusability vs. compliance concerns).

Table C.13
Stated reasons for choosing Helen Keller International (HKI) in open-text form in Study 1.

Theme	Brief description	Example (EN translation; participant ID)
More beneficiaries ("7 > 1") & higher amount (€7 vs. €5)	Preference to help more children with a single donation and/or to send the larger amount.	"With €7 I can help seven children; with nets for €5 I can only help one person." (ID 337)
Concerns about net usage/compliance	Nets protect mainly at night, may be unused/misused/stolen/break; protection not assured.	"…I was also skeptical about how effective a mosquito net is if it only provides protection from bites while you are sleeping…" (ID 77)
Broader health benefits/basic nutrition	Vitamin A strengthens immunity and prevents multiple illnesses (cause-oriented support).	"Supplementing with vitamins can prevent several diseases …" (ID 106)
Cost-effectiveness (lives per €)	HKI perceived to save more lives per euro in the presented setup.	"… the estimated cost-per-life-saved ratio is lower for Helen Keller." (ID 370)
Implementation reliability/ease	Supplement delivery seen as simpler or more reliable than correct net installation/use.	"… A one-time treatment can help. I'm not sure mosquito nets are feasible in recipients' everyday lives." (ID 202)
Other (balancing attention, personal ties, autonomy)	Malaria already well known/funded; desire to back the other cause; personal trust/experience.	"I assumed more people would choose the better-known cause (malaria) and wanted to support the other organization…" (ID 349)

Notes: Translations by the authors; lightly edited for brevity. Multiple themes can co-occur. Participants frequently weighed several considerations simultaneously (e.g., severity/urgency vs. breadth of beneficiaries; durability/reusability vs. compliance concerns).

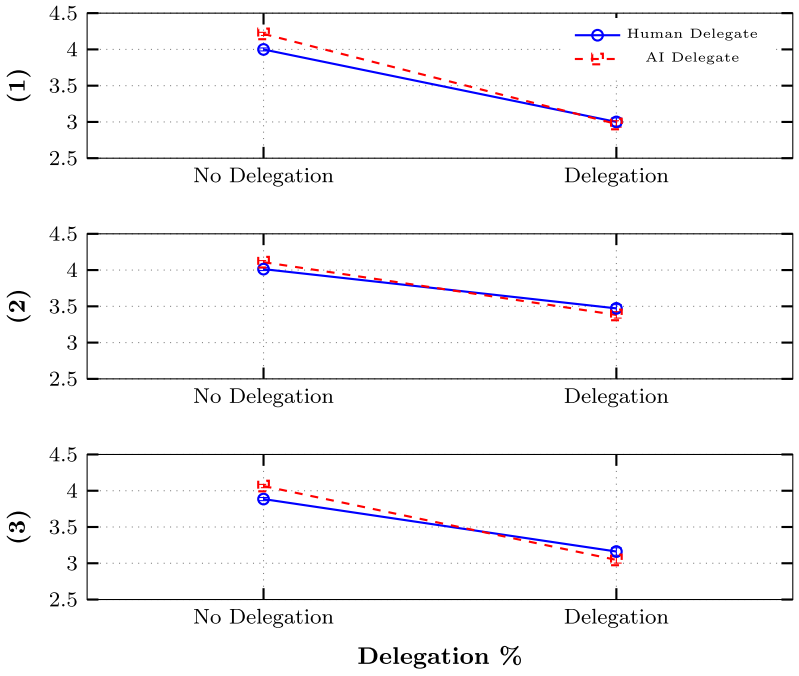


Fig. C.6. Interaction plots for responsibility measures (1)–(3), showing effects of delegation (No vs. Yes) and delegate type (Human vs. AI). (01) "I would like to be fully responsible for the decision, whatever the outcome"., (02) "I feel responsible for the outcome of this decision"., and (03) "I have a moral obligation to make such a decision myself". Error bars represent 95% confidence intervals.

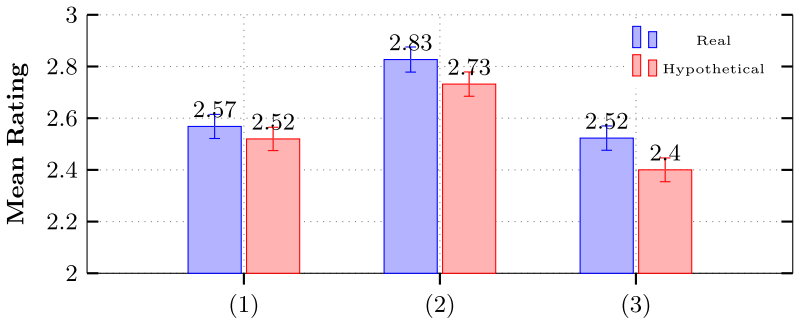
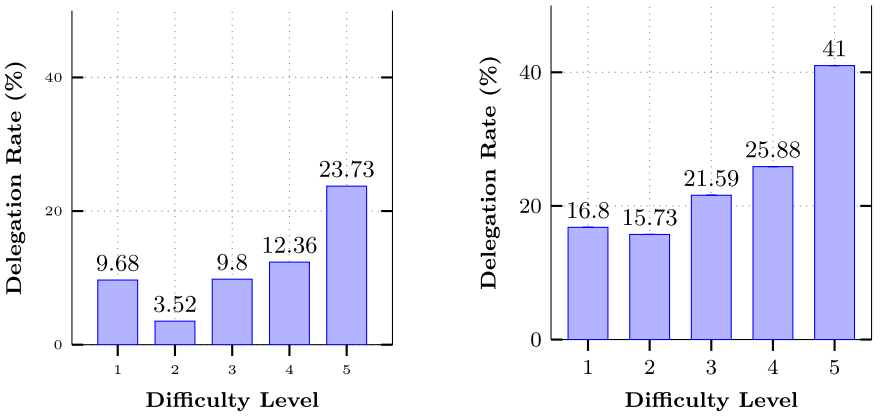


Fig. C.7. Comparison of AI's capability ratings on a five-point scale (1 = strongly disagree to 5 = strongly agree) for three questions: (1) "In a situation as described in this study, an artificial intelligence (AI) can make a better decision between two donations than I can" ($t-test, p = 0.0720$); (2) "I have full confidence that an AI can make a high-quality decision between two donations in a situation like this" ($p = 0.0029$); and (3) "AI can make good moral decisions". ($p = 0.0001$). Error bars represent 95% confidence intervals.



(a) Study 1: Delegation increases as the difficulty level rises from 9.68% to 23.73%

(b) Study 2: Delegation increases as the difficulty level rises from 16.8% to 41.00%

Fig. C.8. Delegation rates for each level of decision difficulty on a 5-point scale for both samples. Difficulty is significantly higher for delegators than non-delegators (Sample 1 $\chi^2(4, N = 800) = 26.16, p < 0.001$, Sample 2 $\chi^2(4, N = 4843) = 136.58, p < 0.001$).

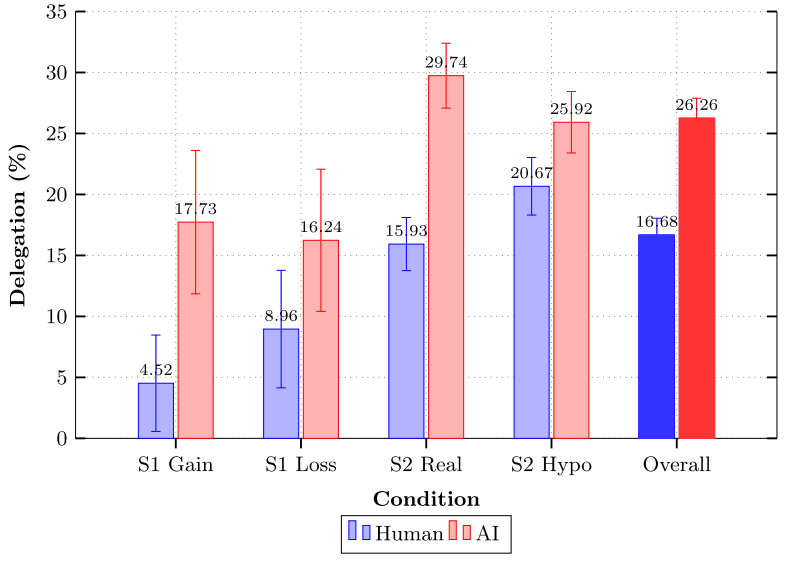


Fig. C.9. Delegation rates by condition for both delegates. Bars show means (%) with 95% confidence intervals. Overall rates are weighted across all situations and both studies.

References

Adam, D., 2024. Lethal AI weapons are here: How can we control them?. Nature 629 (ePub), 521–523. http://dx.doi.org/10.1038/d41586-024-01029-0.

Anderson, M., Anderson, S.L., 2011. Machine Ethics. Cambridge University Press.

Appel, M., Izydorczyk, D., Weber, S., Mara, M., Lischetzke, T., 2020. The uncanny of mind in a machine: Humanoid robots as tools, agents, and experiencers. Comput. Hum. Behav. 102, 274–286. http://dx.doi.org/10.1016/j.chb.2019.07.031.

Argenton, C., Potters, J., Yang, Y., 2023. Receiving credit: On delegation and responsibility. Eur. Econ. Rev. 158, 104522. http://dx.doi.org/10.1016/j.euroecorev.2023.104522.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., Rahwan, I., 2018. The Moral Machine experiment. Nature 563 (7729), 59–64. http://dx.doi.org/10.1038/s41586-018-0637-6.

Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J.B., Shariff, A., Bonnefon, J.-F., Rahwan, I., 2020. Drivers are blamed more than their automated cars when both make mistakes. Nat. Hum. Behav. 4 (2), 134–143. http://dx.doi.org/10.1038/s41562-019-0762-8.

Bartling, B., Engl, F., Weber, R.A., 2014. Does willful ignorance deflect punishment? – An experimental study. Eur. Econ. Rev. 70, 512–524. http://dx.doi.org/10.1016/j.euroecorev.2014.06.016.

Bartling, B., Fehr, E., Özdemir, Y., 2023. Does Market Interaction Erode Moral Values? Rev. Econ. Stat. 105 (1), 226–235. http://dx.doi.org/10.1162/rest_a_01021.

Bartling, B., Fischbacher, U., 2012. Shifting the Blame: On Delegation and Responsibility. Rev. Econ. Stud. 79 (1), 67–87. http://dx.doi.org/10.1093/restud/rdr023.

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

Bonnefon, J.-F., Rahwan, I., Shariff, A., 2024. The Moral Psychology of Artificial Intelligence. Annu. Rev. Psychol. 75, 653–675. http://dx.doi.org/10.1146/annurev-psych-030123-113559.

Bonnefon, J.F., Shariff, A., Rahwan, I., 2016. The social dilemma of autonomous vehicles. Science 352 (6293), 1573–1576. http://dx.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. J. Behav. Decis. Mak. 33 (2), 220–239. http://dx.doi.org/10.1002/bdm.2155.

Castelo, N., Bos, M.W., Lehmann, D.R., 2019. Task-Dependent Algorithm Aversion. J. Mark. Res. 56 (5), 809–825. http://dx.doi.org/10.1177/0022243719851788.

Cath, C., 2018. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 376 (2133), 20180080. http://dx.doi.org/10.1098/rsta.2018.0080.

Chugunova, M., Sele, D., 2022. We and It: An interdisciplinary review of the experimental evidence on how humans interact with machines. J. Behav. Exp. Econ. 99, 101897. http://dx.doi.org/10.1016/j.socec.2022.101897.

Cint GmbH, 2024. Academic Research. https://de.cint.com/academic-research.

Coffman, L.C., 2011. Intermediation reduces punishment (and reward). Am. Econ. J.: Microeconomics 3 (4), 77–106. http://dx.doi.org/10.1257/mic.3.4.77.

Dana, J., Weber, R.A., Kuang, J.X., 2007. Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Econom. Theory 33 (1), 67–80. http://dx.doi.org/10.1007/s00199-006-0153-z.

Danaher, J., 2019. The rise of the robots and the crisis of moral patiency. AI SOCIETY 34 (1), 129–136. http://dx.doi.org/10.1007/s00146-017-0773-9.

Dastin, J., 2022. Amazon scraps secret AI recruiting tool that showed bias against women. In: Ethics of Data and Analytics. Auerbach Publications, pp. 296–299.

Dattner, B., Chamorro-Premuzic, T., Buchband, R., Schettler, L., 2019. The legal and ethical implications of using AI in hiring. Harv. Bus. Rev. 25, 1–7.

Dietvorst, B.J., Bartels, D.M., 2022. Consumers Object to Algorithms Making Morally Relevant Tradeoffs Because of Algorithms' Consequentialist Decision Strategies. J. Consum. Psychol. 32 (3), 406–424. http://dx.doi.org/10.1002/jcpy.1266.

Dietvorst, B.J., Simmons, J.P., Massey, C., 2015. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J. Exp. Psychol. Gen. 144 (1), 114–126. http://dx.doi.org/10.1037/xge0000033.

Dong, M., Bocian, K., 2024. Responsibility gaps and self-interest bias: People attribute moral responsibility to AI for their own but not others' transgressions. J. Exp. Soc. Psychol. 111, 104584. http://dx.doi.org/10.1016/j.jesp.2023.104584.

Dressel, J., Farid, H., 2018. The accuracy, fairness, and limits of predicting recidivism. Sci. Adv. 4 (1), eaao5580. http://dx.doi.org/10.1126/sciadv.aao5580.

Dzindolet, M.T., Pierce, L.G., Beck, H.P., Dawe, L.A., 2002. The Perceived Utility of Human and Automated Aids in a Visual Detection Task. Hum. Factors 44 (1), 79–94. http://dx.doi.org/10.1518/0018720024494856.

Fahrenwaldt, A., tho Pesch, F., Fiedler, S., Baumert, A., 2024. What's moral wiggle room? A theory specification. Judgm. Decis. Mak. 19, e17. http://dx.doi.org/10.1017/jdm.2024.16, Published online by Cambridge University Press: 18 September 2024. Open Access.

Falk, A., Neuber, T., Szech, N., 2020. Diffusion of Being Pivotal and Immoral Outcomes. Rev. Econ. Stud. 87 (5), 2205–2229. http://dx.doi.org/10.1093/restud/rdz064.

Falk, A., Szech, N., 2013. Morals and Markets. Science 340 (6133), 707–711. http://dx.doi.org/10.1126/science.1231566.

Feier, T., Gogoll, J., Uhl, M., 2021. Hiding behind machines: When blame is shifted to artificial agents. ArXiv Preprint. arXiv:2101.11465.

Freisinger, E., Schneider, S., 2024. Decoding decision delegation to artificial intelligence: A mixed-methods study on the preferences of decision-makers and decision-affected in surrogate decision contexts. Eur. Manag. J. http://dx.doi.org/10.1016/j.emj.2024.10.004.

GDPR, 2016. Regulation (EU) 2016/679 of the European parliament and of the council. Regul. (EU) 679, 2016.

Gerke, S., Minssen, T., Cohen, G., 2020. Chapter 12 - ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr, A., Memarzadeh, K. (Eds.), Artificial Intelligence in Healthcare. Academic Press, pp. 295–336. http://dx.doi.org/10.1016/B978-0-12-818438-7.00012-5.

Gert, B., 2005. Morality: Its Nature and Justification. Oxford University Press, http://dx.doi.org/10.1093/0195176898.001.0001.

GiveWell, 2024. Top charities. https://www.givewell.org/charities/top-charities. (Accessed 18 November 2024).

Gogoll, J., Uhl, M., 2018. Rage against the machine: Automation in the moral domain. J. Behav. Exp. Econ. 74, 97–103. http://dx.doi.org/10.1016/j.socec.2018.04.003.

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

Gray, K., Young, L., Waytz, A., 2012. Mind perception is the essence of morality. Psychol. Inq. 23 (2), 101–124. http://dx.doi.org/10.1080/1047840X.2012.651387.

Grossman, Z., van der Weele, J.J., 2017. Self-Image and Willful Ignorance in Social Decisions. J. Eur. Econ. Assoc. 15 (1), 173–217. http://dx.doi.org/10.1093/jeea/jvw001.

Hale, K., 2021. AI bias caused 80% of black mortgage applicants to be denied. Forbes 9, 2021.

Hamman, J.R., Loewenstein, G., Weber, R.A., 2010. Self-interest through delegation: An additional rationale for the principal-agent relationship. Am. Econ. Rev. 100 (4), 1826–1846. http://dx.doi.org/10.1257/aer.100.4.1826.

Holbrook, C., Holman, D., Clingo, J., Wagner, A.R., 2024. Overtrust in AI recommendations about whether or not to kill: Evidence from two human-robot interaction studies. Sci. Rep. 14 (1), 19751. http://dx.doi.org/10.1038/s41598-024-69771-z.

Jago, A.S., 2017. Algorithms and Authenticity. Acad. Manag. Discov. 5 (1), 38–56. http://dx.doi.org/10.5465/amd.2017.0002.

Jussupow, E., Benbasat, I., Heinzl, A., 2020. Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. In: Proceedings of the 28th European Conference on Information Systems. ECIS, pp. 15–17.

Kirchkamp, O., Strobel, C., 2019. Sharing responsibility with a machine. J. Behav. Exp. Econ. 80, 25–33. http://dx.doi.org/10.1016/j.socec.2019.02.010.

Klockmann, V., von Schenk, A., Villeval, M.C., 2022. Artificial intelligence, ethics, and intergenerational responsibility. J. Econ. Behav. Organ. 203, 284–317. http://dx.doi.org/10.1016/j.jebo.2022.09.010.

Köbis, N., Bonnefon, J.-F., Rahwan, I., 2021. Bad machines corrupt good morals. Nat. Hum. Behav. 5 (6), 679–685. http://dx.doi.org/10.1038/s41562-021-01128-2.
Krügel, S., Ostermaier, A., Uhl, M., 2023. Algorithms as partners in crime: A lesson in ethics by design. Comput. Hum. Behav. 138, 107483. http://dx.doi.org/10.1016/j.chb.2022.107483.
Lee, M.K., 2018. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data Soc. 5 (1), 2053951718756684. http://dx.doi.org/10.1177/2053951718756684.
Leiner, D.J., 2019. Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. Surv. Res. Methods 13 (3), 229–248. http://dx.doi.org/10.18148/srm/2019.v13i3.7403.
Leiner, D.J., 2024. Sosci survey (version 3.5.02) [computer software]. https://www.soscisurvey.de.
Longoni, C., Bonezzi, A., Morewedge, C.K., 2019. Resistance to Medical Artificial Intelligence. J. Consum. Res. 46 (4), 629–650. http://dx.doi.org/10.1093/jcr/ucz013.
Mahmud, H., Islam, A.N., Ahmed, S.I., Smolander, K., 2022. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. Technol. Forecast. Soc. Change 175, 121390. http://dx.doi.org/10.1016/j.techfore.2021.121390.
Matthias, A., 2004. The responsibility gap: Ascribing responsibility for the actions of learning automata. Ethics Inf. Technol. 6, 175–183. http://dx.doi.org/10.1007/s10676-004-3422-1.
Metz, C., Satariano, A., 2020. An algorithm that grants freedom, or takes it away. N. Y. Times 6.
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. http://dx.doi.org/10.1126/science.aax2342.
Office of Management and Budget (OMB), 2024. Revisions to omb's statistical policy directive no. 15: Standards for maintaining, collecting, and presenting federal data on race and ethnicity. https://www.federalregister.gov/d/2024-06469. (Accessed 04 December 2024).
Pazzanese, C., 2020. Ethical concerns mount as AI takes bigger decision-making role. https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/.
Rothenhäusler, D., Schweizer, N., Szech, N., 2018. Guilt in voting and public good games. Eur. Econ. Rev. 101, 664–681. http://dx.doi.org/10.1016/j.euroecorev.2017.08.001.
Rudin, C., Wang, C., Coker, B., 2020. The age of secrecy and unfairness in recidivism prediction. Harv. Data Sci. Rev. 2 (1), 1. http://dx.doi.org/10.1162/99608f92.6ed64b30.
Santoni de Sio, F., Mecacci, G., 2021. Four responsibility gaps with artificial intelligence: Why they matter and how to address them. Philos. Technol. 34 (4), 1057–1084. http://dx.doi.org/10.1007/s13347-021-00450-x.
Serra-Garcia, M., Szech, N., 2021. The (In)Elasticity of Moral Ignorance. Manag. Sci. 68 (7), 4815–4834. http://dx.doi.org/10.1287/mnsc.2021.4153.
Shank, D.B., DeSanti, A., Maninger, T., 2019. When are artificial intelligence versus human agents faulted for wrongdoing? Moral attributions after individual and joint decisions. Inf. Commun. Soc. 22 (5), 648–663. http://dx.doi.org/10.1080/1369118X.2019.1568515.
Statistisches Bundesamt (Destatis), 2022. Statistischer bericht: Bevölkerungsfortschreibung zensus 2022. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Publikationen/Downloads-Bevoelkerungsstand/statistischer-bericht-bevoelkerungsfortschreibung-zensus-2022-5124107.xlsx?_blob=publicationFile. (Accessed 04 December 2024).
Steffel, M., Williams, E.F., 2018. Delegating decisions: Recruiting others to make choices we might regret. J. Consum. Res. 44 (5), 1015–1032. http://dx.doi.org/10.1093/jcr/ucx080, arXiv:https://academic.oup.com/jcr/article-pdf/44/5/1015/29013055/ucx080.pdf.
Steffel, M., Williams, E.F., Perrmann-Graham, J., 2016. Passing the buck: Delegating choices to others to avoid responsibility and blame. Organ. Behav. Hum. Decis. Process. 135, 32–44. http://dx.doi.org/10.1016/j.obhdp.2016.04.006.
U.S. Census Bureau, 2023. National population totals and components of change: 2020–2023. https://www.census.gov/data/datasets/time-series/demo/popest/2020s-national-detail.html. (Accessed 04 December 2024).
U.S. Census Bureau, 2024. Comparing race and hispanic origin. https://www.census.gov/topics/population/hispanic-origin/about/comparing-race-and-hispanic-origin.html. (Accessed 04 December 2024).
Vallor, S., 2015. Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character. Philos. Technol. 28 (1), 107–124. http://dx.doi.org/10.1007/s13347-014-0156-9.
Wallach, W., Allen, C., 2008. Moral Machines: Teaching Robots Right from Wrong. Oxford University Press.
Waytz, A., Gray, K., Epley, N., Wegner, D.M., 2010. Causes and consequences of mind perception. Trends Cogn. Sci. 14 (8), 383–388. http://dx.doi.org/10.1016/j.tics.2010.05.006.
Wiese, E., Weis, P.P., Bigman, Y., Kapsaskis, K., Gray, K., 2022. It's a match: Task assignment in human–robot collaboration depends on mind perception. Int. J. Soc. Robot. 14 (1), 141–148. http://dx.doi.org/10.1007/s12369-021-00771-z.
Young, A.D., Monroe, A.E., 2019. Autonomous morals: Inferences of mind predict acceptance of AI behavior in sacrificial moral dilemmas. J. Exp. Soc. Psychol. 85, 103870. http://dx.doi.org/10.1016/j.jesp.2019.103870.
Zhang, Z., Chen, Z., Xu, L., 2022. Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI. J. Exp. Soc. Psychol. 101, 104327. http://dx.doi.org/10.1016/j.jesp.2022.104327.
Zou, L., Khern-am nuai, W., 2023. AI and housing discrimination: the case of mortgage applications. AI Ethics 3 (4), 1271–1281. http://dx.doi.org/10.1007/s43681-022-00234-9.