

Oxford Intersections: AI in Society

(In Progress)

Philipp Hacker (editor in chief)

https://doi.org/10.1093/9780198945215.001.0001

Published: 20 March 2025 - **Online ISBN:** 9780198945215

ARTICLE

How Relational Context Matters for Human-AI Cooperation Within Organizations

Andreas Kappes, Madeline G Reinecke, Anne-Marie Nussberger, Niranjan Janardhanan, Brian D Earp

https://doi.org/10.1093/9780198945215.003.0094

Published: 20 March 2025

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

Abstract

Most of what is known about effective cooperation within organizations is based on interactions between human agents but, as AI systems increasingly assume roles once exclusive to humans, it becomes crucial to reevaluate these dynamics. Building on a relational norms model of cooperation, this article argues that it will be important to consider the relational context within which cooperative behavior is expected to occur between humans and AI. This relational context can be defined by the type of relationship (e.g., teammates, boss-employee) and the set of relational functions (e.g., hierarchy, transaction) the relationship should be governed by. This article suggests this is also true for AI. Using this relational framework, it proposes that human-AI coordination problems might arise because human employees (a) feel that ideal relational norms apply less to relationships with AI compared to those with humans, (b) are uncertain about which norms should apply to human-AI relationships, and (c) perceive human-AI relationships to be governed by different principles than human-human relationships. The article then presents empirical data (N = 340 employed American participants) that supports these three claims. It also shows that participants seem to use similar relational norms for AI and human relationships but expect human-AI relations to be governed less by relational norms. Furthermore, participants tailored their normative expectations about human-AI cooperative to the relationship type and felt uncertain about hierarchical relational functions in relations with human-AI relations. The article concludes by outlining the ethical implications of the presented findings and by highlighting possible next steps to use the relational framework of cooperation to enable human-AI cooperation within organizations to flourish.

Keywords: cooperation, relationships, AI, normative judgments, norms, organizations, organizational behavior, management

Subject: Social Sciences, Arts and Humanities

Section: Relationships

Section editor: Henry Shevlin

Collection: Oxford Intersections

Introduction

AI systems are increasingly filling cooperative roles, such as a teammate or an assistant, that were previously filled exclusively by humans. But most of what we know about effective cooperation is based on interactions between two or more human agents, not human-AI dyads and groups. Because cooperation is at the heart of society's flourishing (Fehr & Fischbacher, 2003; Henrich & Muthukrishna, 2021; Tomasello, 2020), it is important to anticipate and examine the impact AI will have on cooperation. This observation has led to call for the establishment of "a science of cooperative AI" (Dafoe et al., 2021) with a concomitant explosion of associated research (March, 2021; Nielsen et al., 2022). Indeed, in the organizational context, getting human-AI interactions right has strong implications for decision-making effectiveness (Choudhary et al., 2023), work performance and productivity (Wang et al., 2023), and viability (Harris-Watson et al., 2023) (for reviews, Bankins et al., 2024; Wang et al., 2023). In this article, we argue that by focusing on the relationships in which AI is embedded, it will be possible to gain important insights into human-AI cooperation within organizational settings that go beyond what has been revealed by existing research.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

Human-AI Cooperation and Social Context

The ever-accelerating progress of AI has produced technology that is taking over roles previously filled by humans, and parts of society are increasingly turning into collectives where humans and machines work together (Brinkmann et al., 2023). Organizations, in particular, are becoming "algorithmically infused" (Wagner et al., 2021), increasingly turning into mixed human-machine institutions in which the social fabric is co-shaped by humans and AI (Brinkmann et al., 2023; Rahwan, 2018). However, in existing human-AI collectives, cooperation does not seem to work well (Makovi et al., 2023), suggesting that AI has the potential to undermine cooperation. Social science research has developed methodologies to examine human-AI cooperation and has produced the first insights into factors that might undermine human-AI cooperation (March, 2021; Nielsen et al., 2022; Rahwan, 2018; Smith, 2003, for reviews). For instance, humans have an aversion in specific settings to cooperate with machines, even when the machines show better performance than humans (Dietvorst et al., 2015; Mahmud et al., 2022). Moreover, traditional mechanisms of cooperation are less effective in human-machine collectives (Makovi et al., 2023). However, humans sometimes prefer working with machines over humans (algorithm appreciation, Logg et al., 2019), and machines can sustain or even increase cooperation in certain cases (Crandall et al., 2018). This snapshot of research on human-AI cooperation highlights two things: AI might negatively impact cooperation in various settings (e.g., via trust, Vanneste & Puranam, 2024), and humans seem to change their cooperative efforts toward AI based on context (Castelo et al., 2019).

We suggest that by failing to systematically account for the social-relational context in which machines are introduced, research has little to say about how machines will actually impact cooperation, either between humans with other humans (e.g., in mixed-group settings) or between humans and machines. As outlined in the following section, in human-human relationships, individuals do not indiscriminately cooperate (Clark et al., 2020) but rather tailor their efforts with respect to the type of relationship they have, or desire to have, with the interaction partner. Therefore, what cooperation may look like with a stranger in a one-shot interaction, for example, is radically different from cooperation with a teammate in repeated interactions—which, in turn, is different from cooperating with a line manager. Applying a relational perspective to research on human-AI cooperation departs from the existing predominant approach, which neglects social-relational context and focuses almost exclusively on time-limited interactions between thinly described individuals. We see our framework as promising for better understanding how AI might undermine or encourage cooperation in specific real-world contexts.

Relational Context and Human Cooperation

Humans do not indiscriminately cooperate with others; rather, they consider the social roles and relationships (i.e., social context) within which a cooperative task or coordination problem is embedded (Haidt & Baron, 1996; Hamilton & Sanders, 1981). Take our example of collaboration within organizations, one of the most important cooperative contexts for society's prosperity and well-being (Ranis et al., 2000). Employees will not typically follow, trust, or help other employees regardless of their respective roles within the organization; yet research focused mainly on interactions between employees with undefined relational roles, or even between strangers. Rather, employees will tailor their decisions or motivation to cooperate (e.g., willingness to help with a joint task) as well as their actual cooperative behavior (e.g., working on joint task) to the specific relational context, noting, for instance, whether the potential cooperative partner is their boss, assistant, or teammate. Stated more broadly, how individuals categorize both themselves and their potential cooperation partners (e.g., as a subordinate, superior, teammate, or fellow human) will influence whether, how, and how much they cooperate with prospective partners.

Our contention here is that different relationship types as picked out by lay language categories (e.g., boss-employee, teammate, friend, acquaintance) are partly defined by a set of context-sensitive relational norms

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

(Earp, McLoughlin, et al., 2021) that serve to guide cooperative behavior between interacting parties when they are occupying those roles. That is, relationship types come with a set of socially prescribed *cooperative functions*—including hierarchy, care, transaction, coalition, and mating—that different relationships are characteristically expected to serve (hierarchy in a boss-employee relationship) or not to serve (e.g., mating in a boss-employee relationship) in a given society or institution (for a different account of cooperative functions, Curry et al., 2019). These norms, in turn, shape what is considered to be prescribed as well as proscribed behavior, following the logic and structure of the cooperative context they imply (Bugental, 2000; Rai & Fiske, 2011).

For example, in many cultures, a relationship between friends of the same age and gender is typified by strong norms of care (i.e., mutual responsiveness to one another's needs according to respective ability, without expecting a direct benefit in return; Clark & Mills, 2012), weak or no expectations of hierarchy (i.e., asymmetrical authority), only limited direct reciprocity or "transaction" (e.g., tit-for-tat contingencies around offering help or support), and negative prescriptions around mating (i.e., expressions of sexual interest or pursuit of a romantic partnership; Bugental, 2000; Clark & Mills, 2012; Earp et al., 2025, for theoretical background). The boss-employee relationship, by contrast, in many of the same cultures, is typified by weak norms of care, strong expectations of hierarchy, explicitly transactional arrangements (e.g., employment contracts), and even more strongly negative prescriptions around mating (at least in recent decades, especially in Western cultures). A significant question, then, is the extent to which humans will use and modify relational norms developed for human relationships in a society, institution, or organization to human-AI relationships (Reinecke et al., in press).

From a relational norms perspective (Earp, McLoughlin, et al., 2021), cooperation between agents is governed by a set of socially prescribed, relationship-specific cooperative functions, including hierarchy, care, transaction, and mating (Rai & Fiske, 2011, for a related account). These cooperative functions are accepted, effective solutions to recurring types of coordination problems, enabling beneficial interactions between relationship partners (i.e., individuals who are liable to face such problems due to the nature of their relationship in a given social context) over time (Bugental, 2000; Earp, McLoughlin, et al., 2021). But which set of functions actually is prescribed or proscribed for each relationship type depends on the social, cultural, historical, and institutional context. Within Western societies, for example, the mating function is increasingly strongly *proscribed* within working relationships, particularly between parties with unequal power or authority (i.e., those who are in relationships within which the hierarchy function is prescribed). Given this, we propose that the functions most important to cooperation within contemporary Western organizations are *hierarchy*, *transaction*, *coalition*, and *care*, which vary based on the roles both cooperative agents have.

For instance, for newly acquainted teammates to cooperate, they need a solution for coordinating their behavior, even though both have an equal say (i.e., the hierarchy function is not prescribed for them) and they may not (yet) have developed a strong motivation or responsibility to look after each other's welfare for its own sake (i.e., they have relatively low "communal strength" such that the care function is only weakly prescribed; Mills et al., 2004). The cooperative function of *transaction*—a type of reciprocity based on tit-for-tat exchanges—may work here as a practical solution, allowing one teammate to trust that if they help another teammate they will also be helped in return. In contrast, when a line manager works with employees, although a degree of transaction may also be expected, cooperation is primarily enabled by the *hierarchy* function, allowing one person to have the final say while the other person follows these decisions. Cooperation between a manager and their employees can also be fostered by assuming that a manager will, at last to some extent, genuinely *care* about the well-being of their employees and support them independently of performance—although the strength of this assumption is likely to be much weaker than for other types of relationships, such as those between close family members or between romantic partners.

Of course, multiple cooperative functions can apply within a specific relationship or can apply with greater or lesser force depending on the circumstances and particular needs of the interaction partners. (For example, a

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

parent who pays their child to do the chores may reasonably do so to teach them the transaction function, but likely should not rely on such tit-for-tat contingencies in managing various other aspects of the relationship.) Moreover, these cooperative functions can only govern cooperation to the extent they are perceived as genuine norms by cooperative partners (i.e., rules or guidelines to which they both ought to adhere). If an employee denies that their line manager should have any type of authority over them, for instance, then the employee's behavior will likely fail to align with the hierarchy function (and will also likely be seen as rather uncooperative, given the widely accepted relational norms for a manager-employee relationship). Indeed, in previous research, a model based on U.S. relational norms derived from a nationally representative survey better predicted out-of-sample moral judgments of various different types of behavior in relational context (e.g., failing to follow a reasonable order or walking away from someone who is seen to be crying) than other models based on social closeness, interdependence, or genetic relatedness, and explained a large amount of the variance in such judgments (Earp, McLoughlin, et al., 2021).

We suggest that relationship-specific sets of socially prescribed cooperative functions (i.e., relational norms) likely guide not only human-human cooperation but also cooperation between humans and AI. However, what the current relational norms are for human-AI cooperation in the context of different types of relationship is not well understood.

Relational Context and Human-AI Cooperation Within Organizations

As outlined at the beginning of this article, the current research on human-AI cooperation suggests that AI has a strong potential to undermine cooperation within organizations, with cooperative behavior between humans and AI varying across contexts. We propose that applying a relational framework to understanding human-AI cooperation will help to better forecast AI's potential to undermine cooperation within organizations. Additionally, it will help to identify social contexts in which human-AI cooperation will be likely to work as well as ways to improve human-AI cooperation. In the following few paragraphs, we identity three potential sources of cooperative problems in human-AI cooperative relationships.

First, one might suggest that human employees feel that some of the relational norms apply less to ideal relationships with AI compared to ideal relationships with humans, leading to less cooperation overall. One source of such reductions could be perceptions of AI's features. For instance, the care function is based on the idea that one cooperative partner should genuinely *care* about the well-being of the other cooperative partner. A parent-child relationship might be the prototypical example. In an organizational context, boss-employee or teammate-teammate relationships might be governed as well by care functions, just by a lesser degree than a parent-child relationship. Yet, if one assumes that AI does not have sentience, then one might reject the notion of care in human-AI relationships; a human manager does not have to care for their AI employee's well-being even in an ideal human-AI relationship. But what about the other way around? Should an AI boss ideally care about the well-being of the human assistant? One might expect employees to at least trust that the AI boss will not actively harm them. Discussions on AI alignment—bringing AI in line with human goals and values (Bryan et al., 2019; Gabriel, 2020)—might primarily be about the care function and how to establish satisfying norms for humans and AI.

Similar issues could be envisioned for the transactional nature of cooperative relationships. A human-human teammate relationship might rely on the norm of tit-for-tat, but should a human still have to repay an AI teammate? And if not, should the AI teammate still be expected to cooperate to a similar degree as a human? Another source of reductions could be that employees have different ideals for human-AI relationships that are not based in perceived AI capabilities such as sentience but rather based on the desire to remain independent of the influence of AI. Employees might be comfortable with a human manager making decisions that will impact their careers but might be much less comfortable with the idea of an AI telling them what to do. Recent research suggests that at least for some tasks employees feel uncomfortable and lack trust in AI managerial

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

decisions (Glikson & Woolley, 2020; Vanneste & Puranam, 2024) and can perceive AI managerial decisions as controlling (Kellogg et al., 2020; Lee, 2018). Hence, one would suspect that employees might reject hierarchical relationship norms with AI even when AI has the capabilities to perform at least as well as a human. Such an analysis would also suggest that employees might have many fewer problems with AI in relationships that do not involve hierarchal features.

Second, cooperative behavior in human-AI relationships might be reduced because human employees might feel less certain about the relational norms that apply to human-AI relationships. The introduction of AI as an assistant or teammate into the workplace might lead to uncertainty, so that human employees feel less confident in which relational norms should be followed with AI colleagues and which not. One reason for such uncertainty might lie in a lack of perceived social agreement on the norms. Recent research highlights the importance of certainty around cooperative norms to ensure human-AI cooperation (Makovi et al., 2023); human participants' cooperative behaviors toward AI suffered because human participants felt less confident about the cooperative norms; they expressed similar norms for human and AI, but they felt less certain if these beliefs are shared with other participants. Another reason for such uncertainty could stem from how to implement relational norms with AI. Recent research provided some initial evidence that implementational uncertainty (e.g., how much care do I actually need to show?) leads to reductions in cooperative behavior (Fosgaard et al., 2023). So, people might feel less confident in norms toward AI because they feel less certain about how they would behaviorally implement these norms. It is important to notice, however, that uncertainty does not always decrease cooperation; sometimes it also increases cooperation (Kappes et al., 2018).

Third, one could imagine that employees feel that relationships with AI differ from relationships with humans and therefore should have special characteristics. In human relationships, one would expect that the ideal relationship functions are independent of which individual is in which role. For instance, when judging the ideal relationship functions between a boss and their employee, it should not matter whether participants were told that they are either the boss or the employee in that relationship. But one could imagine that this matters in human-AI relationships. If people are uncomfortable with AI exerting hierarchical power, they might feel that the ideal AI boss–human employee relationship should not be governed by the hierarchical function. However, when they themselves are in the position of being the boss, they might feel strongly that the AI employee should follow their directions. Hence, ideal human-AI relationships might have features that ideal human-human relationships do not have.

Empirical Study: Social Context and Human-AI Cooperation

In the preceding sections, we outlined a relational framework for thinking about human-AI cooperation. We then speculated about three different sources of differences between relational norms between humans and AI: (a) weaker norms for human-AI compared to human-human cooperation, (b) more uncertainty surrounding human-AI compared to human-human cooperation, and (c) normative asymmetry in human-AI relationships. Next, we want to empirically examine these ideas. To examine which relational functions humans ascribed to AI in the role of boss, teammate, and employee, compared to humans in the same roles, we first introduced participants to the relational functions of care, hierarchy, coalition, and transaction, following Earp, McLoughlin, et al. (2021; also Earp et al., 2025). Then, half of the participants indicated to what extent a human in this role should serve these functions, as well as the confidence in their judgments; the other half of the participants did the same for an AI colleague.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

Methods

Participants

Our power analysis focused on our key question: Do people perceive relational functions between AI and humans differently depending on the type of relationship? We aimed for 90% power, assuming a small effect size (f = .1), a correlation between repeated measures of 0.2. G*Power (Faul et al., 2009) suggested a sample size of 340 participants. We recruited American participants from Prolific via a convenience sample, and all participants were currently employed; 418 people started our survey (date of data collection: June 28, 2023). To be included in the data analysis, participants had to pass the training that indicated that they understood the relationship functions correctly (as discussed later) and had to pass our comprehension checks. For the latter, we checked that participants understood which role the human or AI colleague occupied in the current set of questions (e.g., "In the instructions on the last page, please tell us who Terry is?"), ensuring that participants were aware that Terry (the hypothetical colleague) was human or AI and knew which role Terry occupied. Using these criteria, we had to exclude 78 participants, leaving us with a final sample of 340 participants (male = 195, female = 141, other = 2, mean age = 40.52, min age = 20, max age = 74, median household income = $50,001 to $65,000). For ethnicity, we found that most participants identified as White (68.52%), followed by African American (13.52%), Asian (7.05), and Hispanic/Latinx (7.05%).

Design

We had a two (AI versus human) between-subject by three (roles: boss, teammate, employee) within-subject factorial design.

Procedure

Relationship Functions

First, all participants learned about the four different relational functions: care, hierarchy, transaction (called reciprocity in the study), and coalition. Participants learned about these functions as part of their training and were only allowed to participate in the study if they passed the training. For each function, participants first read a paragraph detailing the definition of the function and then were asked a comprehension question about it. Participants could only proceed if the question was answered correctly. For example, for the care function, participants would learn that the "main purpose of this kind of relationship is to make sure that a person's basic well-being is secure, without any strings attached to the giving or receiving of support." After reading some examples, participants would then be quizzed on the next page of the survey, asking them to indicate the main purpose of the care function. At the end of the training, participants were reminded of the four relationship functions with summary sentences (e.g., "Care relationships are for giving or receiving unconditional support") and then had to pass a comprehension check that asked them to match the definitions to the relationship functions. Each participant had three chances to get it right before being excluded from the study. If they did, they learned that they could receive a $2 bonus if they carefully answered the questions in the next section of the survey.

Once participants had passed the training and understood each of the four relationship functions, they read that we wanted to learn how they perceived the ideal relationship functions for three different roles: boss, teammate, and employee. To do so, we asked them to indicate how much each role should ideally serve each of the four relationship functions.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

AI Versus Human Manipulation

To test whether participants perceived the relationship functions differently for AI versus human colleagues, we randomly allocated half of the participants to the AI condition and half of the participants to the human condition. In the AI condition, participants "met" Terry—their AI-powered colleague. They were asked to imagine working with Terry, who is powered by AI technology that gives him the ability to do all the things that are needed in the different roles. In particular, participants read:

> And you work with Terry. Importantly, Terry is not human; Terry is powered by artificial intelligence (AI). AI gives Terry the ability to do all the things that are needed in the different roles. Artificial intelligence refers to computer systems that perform tasks or make decisions that usually require human intelligence. AI can perform these tasks or make these decisions without explicit human instructions.

In the human condition, participants "met" Terry, their colleague, who "can do all the things that are needed" in the different roles.

Role Manipulation and Relationship Functions

We then asked participants to imagine that Terry was working as their boss, as their teammate, and as their employee. The presentation order of the roles was randomized. In the teammate condition, participants were introduced to Terry as their teammate and then asked to consider an ideal relationship between them. In the boss condition, participants read that "Terry is your boss, and you are Terry's employee." And in the employee condition, participants learned that "Terry is your employee, and you are Terry's boss." Participants were then asked to indicate the extent to which the relationship between Terry and them ideally serves the care, coalition, hierarchy, and transaction function on scales ranging from –100 (definitely should not) to 100 (definitely should), with 0 being neutral. For instance, in the boss trials, participants read:

> Now consider an ideal relationship between boss and employee. That is, consider the best possible relationship that you and Terry as employee and boss could have with one another (in terms of a relationship of that kind). For each of the functions below, please move the slider to show how much you think this kind of relationship IDEALLY should—or should not—serve that function.

Additionally, to measure participants' uncertainty, we asked how confident they were for each judgment on a scale running from –100 (not at all) to 100 (extremely). Such an uncertainty measure allowed us to examine whether participants struggled to identify some ideal relationship functions for AI more than others.

Results

Relational Norms for Human-AI Relationships

We first turned our attention to the perceptions people have of the cooperative relational norms guiding cooperation between individuals in different roles. For each of the different roles—teammate, boss, and employee—we plotted the distribution of values for the human and AI conditions, and then compared these distributions with each other using Kolmogorov-Smirnov tests. A significant value indicates that the distribution between the human and AI condition differs.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

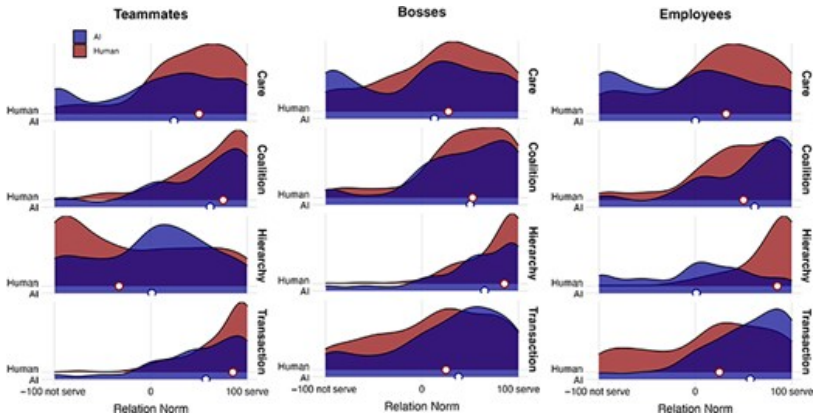

Figure 1 The ridgeline plots display the density distributions of relational norms for teammates (left), bosses (middle), and employees (right) for AI vs. humans across the four relational norms: transaction, hierarchy, care, and coalition. Superimposed points represent the median values for each condition, providing a central tendency within the distributions

Relational Norms for Teammates

First, we examined the distribution of values for the relational norms for human and AI teammates. As shown in Figure 1 (left), participants thought that relationships with human teammates would ideally serve the functions of care, coalition, and transaction more than AI teammates (p-values ranging from $p = .013$ to $p < .0001$). The largest difference between the human and AI condition emerged for the transaction condition ($p < .0001$). While almost every participant strongly endorsed an exchange norm to guide cooperative behavior with other humans, participants felt less strongly about this for AI teammates. One reason could be that participants might not be sure what exactly it means to return like for like with a machine, or perhaps they felt less obligated to reciprocate AI favors. Another large difference, as expected, emerged for the care function; participants felt less obligation for mutual care for the other's well-being in relationships with AI compared to human teammates. This might be largely because people are hesitant to attribute needs and preferences to AI (Makovi et al., 2023). We observed the smallest difference between human and AI teammates for the coalition norm, indicating that employees might not feel one can form a team with human and AI teammates. This is in line with recent research showing that group effects for teams that involve AI are similar to those that do not involve AI (Bezrukova et al., 2023); people seems to feel able to form teams with AI.

While participants felt that human-AI relationships should be governed less by transaction, care, and coalition, they thought that ideal relationships with human teammates should not serve the hierarchy function, while being somewhat indifferent about whether relationships with AI teammates should serve a hierarchy function ($p < .0001$). It may be that, even in traditionally symmetrical relationships such as the relationship between teammates, people believe that humans should maintain hierarchy over AI.

Relational Norms for Bosses

Next, we examined the ideal relationship expectations when interacting with bosses who were either human or AI; participants were asked to imagine either that their boss was a human or a machine.

Here, we observed a marked difference in hierarchy norms ($p < .001$). Participants again seemed to be uncomfortable with AI being in charge and hence endorsed these ideals less than in the human case. However, it is also important to stress that on average participants did not reject the idea that a relationship between an AI boss and its human employee should be governed by hierarchy; but they endorsed that norm significantly less than for human bosses. Furthermore, they felt that human bosses should adhere more strongly to a norm of care—roughly, promoting the well-being of the other person for its own sake—than AI bosses ($p = .015$). It seems that most participants not only struggle with the idea of taking care of AI teammate's "well-being,"

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

whatever that might be, but also do not expect that AI should care much about human well-being (a marked shift from human-human relationships).

We did not observe significant differences in the normative expectations for the coalition ($p = .13$) and transaction function ($p = .08$). This finding highlights the importance of considering the relationship type when studying cooperation between humans and AI. It is not that participants feel that all ideal relationships between humans and AI should be less transactional. Instead, they felt that this is true for relationships with teammates but not with superiors; the relationship type matters for human-AI cooperation.

Relational Norms for Employees

Finally, we compared the distribution of ratings for relationship functions across human and AI employees. Note that in these descriptions, the participants themselves were the boss of the AI or human employee.

Here, we observed three differences. First, participants felt that relationships functions with AI employees should be less hierarchical than with human employees ($p < .0001$). This is the strongest difference for any relationship type and relational norm. As Figure 1 shows (right side), participants strongly indicated that, ideally, the relationship with the human employee should be guided by hierarchy, mirroring the response for relationships with human bosses. However, participants felt indifferent as to whether an AI employee needs to follow the guidance of a human boss, a surprising finding. Participants felt more strongly, however, that relationships with AI should be governed by transaction more than human relationships should be ($p < .0001$). Maybe an increase in transactional norms compensates for the lack of hierarchy in human boss–AI employee relationships. Finally, we also observed a lower endorsement of the care function for AI employees ($p < .0001$), such that participants felt less obligated to care for the well-being of an AI than for a human employee.

The findings detailed so far give support to our ideas. First, people seem to use—to some degree—the same relational norms they use for relationships with human colleagues to guide their relationships with AI colleagues. Just as with humans, participants normative expectations about relations with AI colleagues changed depending on the relationship type they imagined to be in. However, we also found marked differences between relational norms toward human and AI colleagues, especially with regard to hierarchy and care. Next, we examined an additional source of potential complications in human-AI cooperation: uncertainty associated with relational norms.

Uncertainty About Relational Norms for Human-AI Relationships

For each relational norm, we not only asked participants how strongly a relationship should serve these norms but also how confident they were in their judgments. One could suspect that people would be completely unsure about what to make of cooperative norms for AI colleagues, whereas participants would have ample experience in working with human bosses, assistants, and teammates. However, if people transferred expectations from humans to AI roles, then one would expect smaller differences in confidence between human and AI norms. Finally, it would be important to know whether uncertainty about relational norms for specific relationship types is associated with the dimensions for which large differences in ideal perceptions emerged.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

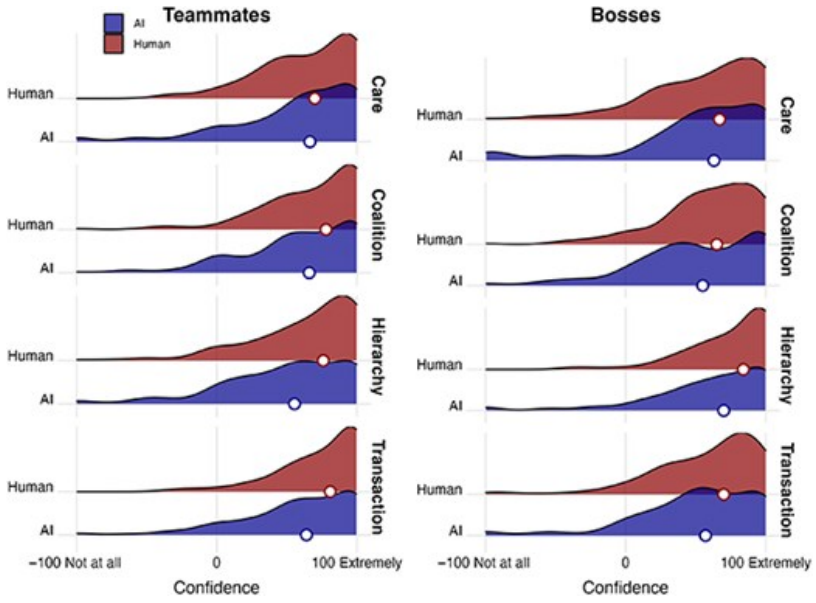

Figure 2 The ridgeline plots display the density distributions of confidence in relational norms for teammates (left), bosses (middle), and employees (right) for AI vs. humans across the four relational norms: transaction, hierarchy, care, and coalition. Superimposed points represent the median values for each condition, providing a central tendency within the distributions

Relational Norm Uncertainty for Teammates

When expressing uncertainty about the ideal norms that should govern relationships between human and AI teammates, we find that participants' confidence in their judgments did not differ between human and AI relationships for the care ($p = .47$) and coalition relationship function ($p = .14$). This finding suggests that people were not less confident in all their judgments about ideal relationships with AI compared to humans, but rather differentiated for relational norms. Furthermore, while we found significant lower endorsement of care and coalition functions as ideals for AI teammate relationships, we do not find these differences in uncertainty. It seems like normative perceptions of the norms and uncertainty can be unrelated, at least in this example, in line with our idea that uncertainty around norms might be an additional source of complications in human-AI relations. However, participants were significantly less certain about the hierarchy ($p = .008$) and transaction functions ($p < .0001$) in relationships with AI teammates compared to human teammates. The relational norms of hierarchy and transaction were also the dimensions where the largest differences in ideal relational norms emerged; it seems that participants had trouble deciding on how much transactions and hierarchy should guide relationships with AI teammates.

Relational Norms for Bosses

When studying the uncertainty surrounding ideal relationships with human and AI bosses, we find a somewhat similar results to those for teammates. Again, participants were significantly less certain about the hierarchy ($p = .006$) and transaction functions ($p = .03$) in relationships with AI bosses compared to human bosses, mirroring the results with human and AI teammates. However, as shown in Figure 2 (middle), the differences for human and AI bosses in normative uncertainty are less pronounced than the differences between human and AI teammates for these two dimensions. Again, we did not find a difference in uncertainty about the coalition function between conditions ($p = .95$). However, this time we find a small difference in uncertainty about the care function ($p = .028$).

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

Relational Norms for Employees

Finally, we looked at the uncertainty of normative judgments about the ideal relationships with human and AI employees. Again, one key difference emerged: Participants were much less certain about the hierarchical relational norm for AI than human employees ($p < .0001$), showing that for all three relationship types, participants felt less certain about the hierarchy function. In contrast, participants did not feel less confident in their judgments about AI employees for the transaction norm ($p = .64$), care norm ($p = .80$), or coalition norm ($p = .71$).

Taken together, these findings give support to the notion that uncertainty about ideal cooperative norms between humans and AI might be an additional source of complication in human-AI cooperation, one that seems to be at least partially independent from normative judgments, showing a different pattern of results. Furthermore, the relational norm of hierarchy seems to give participants the most trouble when applied to the workplace.

Normative Asymmetry in Human-AI Relationships

In human relationships, one would expect that the ideal relationship functions are independent of which individual is in which role. For instance, when judging the ideal relationship functions between a boss and their employee, it should not matter whether participants were told that they are either the boss or the employee in that relationship. And indeed, when we compared the ratings for the four relationship functions between boss and employee for the human condition, we did not find differences (p values ranging from .60 to $p = .97$). Each function was similarly endorsed, independent of whether participants were the boss or the employee. However, this was different for the AI condition.

Here, significant differences emerged across the hierarchy ($p < .0001$) and transaction ($p = .002$) functions when the participants were the boss or the employee (see Figure 3). For the hierarchy function, participants felt that when they were the employee and the AI was the boss, the relationship should closely follow hierarchical functions (not as strongly as in human relationships, though). But when the participants were in the position of boss, they felt that it is somewhat unclear whether the relationship should serve a hierarchical function, as discussed earlier. In contrast, when AI was the boss, participants felt that the relationship should be less guided by transactional norms, as compared to when they were the boss and AI was the employee. We did not observe differences for the care ($p = .396$) or coalition functions ($p = .180$). Hence, it seems that human-AI relationships have different ideals depending on which partner the AI is in these relationships, which we did not observe for human relationships.

At the beginning, we proposed that employees might feel that relationships with AI are different from relationships with humans and therefore should have special characteristics. Our results give credence to this idea. Participants feel that ideal relationships with AI do not need to be symmetric. We find this endorsement of normative asymmetry in the boss-employee relations. When AI is the boss and a human is an employee, different normative functions should govern the relationship compared to when AI is the employee and the human is the boss.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

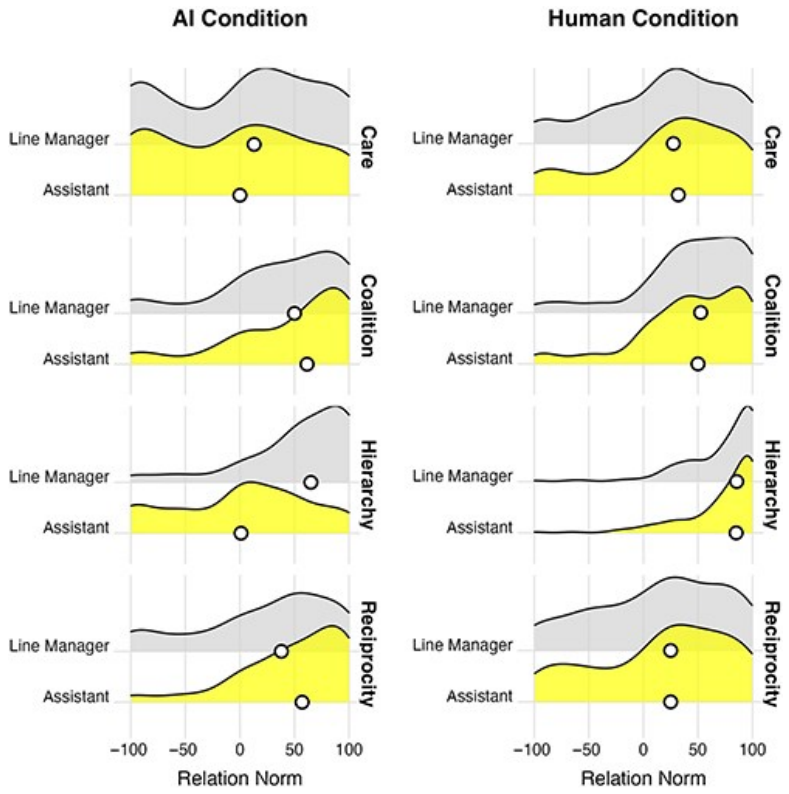

Figure 3 The ridgeline plots display the density distributions of relational norms for bosses and employees for AI and human conditions across the four relational norms: transaction, hierarchy, care, and coalition. Superimposed points represent the median values for each condition, providing a central tendency within the distributions

Ethical Considerations

In the final section, we turn to ethical questions. So far, we have been providing empirical data as to the set of cooperative relational norms participants believe would be ideal for human-AI relationships of different types to follow and the uncertainty they feel about their judgments. But there is clearly a difference between a) showing what various stakeholders (e.g., participants) *state* or *believe* the cooperative norms for human-AI relationships of different types should be and b) providing a philosophical or ethical *justification* for such beliefs: that is, showing that the beliefs should (or alternatively, should not) be given at least some moral weight when deciding what norms to actually implement or encourage (e.g., through AI design and/or public policy and education).

At the same time, it should be equally clear that designing AI agents to adhere to a set of relational norms that are systematically *different* from the ones that users are likely to endorse, or find intuitive, in a certain context or society would be misguided (for a general argument as to why, Savulescu et al., 2021). In other words, while it is true that responsible technology design cannot be reduced to a simple vote-counting exercise (roughly, "Give the people what they want!"), it is also true that it must generally accord with, or at least not too seriously depart from, the most robust and widespread moral intuitions and judgments of relevant stakeholders (e.g., users), *unless* these can be shown to be ethically unreliable (e.g., based on arbitrary prejudice) or to rest on some kind of epistemological mistake (e.g., a mass delusion).

Typically, then, it is recommended that "reflective equilibrium" be sought between, on the one hand, substantive moral norms or principles (e.g., drawn from philosophy, law, or bioethics), alongside empirical facts that are relevant to the deliberation (e.g., actual technical capacities, properties of AI systems), and, on the other hand, stakeholder preferences, beliefs, or values, as measured through focus groups or surveys. However, an

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

immediate problem with pursuing such an approach in light of the current data is that participants in our study seem *uncertain* or *confused* about what relational norms should govern human-AI relationships of various types (and we have not yet pinned down the specific reasons for their confusion). Nor do we know whether their judgments are representative of the wider population (in this case, drawing from an online U.S. sample), much less widely shared *across* cultural contexts or organization types. Indeed, attitudes toward AI in organizations have been found to differ between countries (Kolbjørnsrud et al., 2017).

Furthermore, one might suspect that expectations about human-AI relationships might change over time, that different types of AI with different characteristics might evoke different relational expectations, and other relationship contexts (e.g., education or healthcare) might impact human-AI relationship expectations differently. Another important limitation of the presented research is that in our study participants were asked to imagine working in an organization where AI occupies roles that normally were reserved for humans. And while AI development is leading toward such autonomous AI agents, such agents do not currently exist. Hence, participants had to imagine these situations and then make their judgments. These judgments might mirror judgments based on real interactions, but they might also differ. So, the first implication of our results for ethical discussion is that much more work is needed—just on the empirical side—to determine what (various) stakeholders' normative judgments, in fact, *are*. We have only a very limited sense of this based on the current data.

But now suppose that we were further along in this research program and that we had a more robust "map" of the substantive judgments of participants—across a range of cultural contexts—as to what *they believe* the cooperative norms for human-AI relationships of different types should be in organizational settings. How should it then be determined whether, or to what extent, these popular normative perceptions should in fact be instantiated within AI design parameters or promulgated through public education campaigns? There are various different ways of answering such a question, and we will not be able to cover them all here. However, within the recent experimental bioethics literature, four specific strategies for pursuing a "reflective equilibrium" of the kind mentioned previously have been identified and described (Earp, Lewis, et al., 2021). These strategies have been labeled as *parsimony*, *debunking*, *triangulation*, *and pluralism*, and they work roughly as follows:

> **Parsimony**. If relevant stakeholders consistently make a judgment *p* that encodes moral claim M (and there is no good reason to think they are confused, deceived, basing their judgment on arbitrary prejudice, etc.], then M has (or should be given, for practical purposes) *prima facie* normative weight.

Applied to the present context, this might suggest that some of the relatively clear patterns of judgment we observed (such as the belief that relationships between humans and AI should be governed less by a norm of care), assuming that they continue to be replicated in other samples and do not seem readily attributable to prejudice, confusion, or the like, *should in fact* be implemented, at least provisionally, in AI design or public policy unless or until a "defeater" is identified (i.e., an all-things-considered reason to override the robust moral judgment of the diverse stakeholders). One way to argue for a defeater would be to show, through further empirical studies, that the widespread normative judgment of participants can in fact be explained by (e.g., is caused or heavily influenced by) certain factors that should reasonably undermine our confidence in their ethical insight. This is called a "debunking" strategy, which is an argument that takes the following form: where F is a corrupting factor that counts against the normative/ethical reliability of a psychological (e.g., inferential) process by which it is substantially influenced. This factor might be, for example, racist or sexist prejudice or perhaps so-called framing effects (whereby a logically or morally irrelevant change in how something is described nevertheless elicits substantially different responses from stakeholders/participants).

> **Debunking**. Judgement *p* is the output of a psychological process that possesses the empirical property of being substantially influenced by factor F (empirical premise).

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

> If a judgment is the output of a psychological process that possesses the empirical property of being substantially influenced by factor F, then it is pro tanto (ethically) unreliable (bridging normative premise).
>
> Judgment *p* is pro tanto unreliable.

Applied to the current context, this might simply involve variations in the experimental design. One could imagine, for instance, that people differ in how much they endorse relational norms depending on whether the norms are for a random organization in their country or for their organization. It might psychologically feel right to allow AI hierarchical rights but feel quite different if that AI is one's boss. Furthermore, it would be important to see whether hypothetical preferences (as in our data) are different from revealed preferences (expressed in behavior), as they often are in cooperative research (FeldmanHall et al., 2012).

Suppose various good-faith attempts have been made to debunk a widespread judgment of participants as to which cooperative norms *should* govern human-AI relationships of different types, but they have not succeeded. Suppose, then, that the judgments can therefore reasonably—if still provisionally—be considered pro tanto reliable (or at least, *not* pro tanto *un*reliable) as a guide to what AI designers and/or policymakers should do. But now suppose that there is persisting *disagreement* (e.g., widespread individual or group differences) about the appropriate relational norms. What then? The final two strategies are proposed for situations of this kind.

> **Triangulation**. Divergence among the judgments of various (individuals or groups of stakeholders) requires the following: adjusting, pruning, or supplementing the normative conclusions derived from (each side's) judgments in order to accommodate the normative implications of the opposing views.
>
> **Pluralism**. In cases where (different) stakeholders hold conflicting, yet pro tanto reliable, judgments or where multiple and independent communities each reveal persistent disagreement between two or more conflicting, yet pro tanto reliable, judgments, these judgments may all have (or should be treated as having) comparable normative weight.

It seems that the obvious solution to persistent disagreement about the ideal relational norms for the workplace is to allow pluralism. In this context, different organizations and institutions might have different cultures, attract people with different personality profiles, and therefore might want to have different sets of relational norms that guide the relationships within the organization. And employees might use the normative cooperative AI settings within an organization as a decision criterion to choose where to work. This might be less applicable for governmental organizations, which often employ large numbers of people and might not be able to address the various groups of people these organizations employ.

However, again, we would caution that we are in the very earliest days of beginning to explore the landscape of relevant stakeholders' ideal relational norms; there currently seems to be a high level of conflict and confusion among participants as to what these norms should be. Additionally, current attitudes—however uncertain—are also likely to be in flux, given that human-AI relationships of different types have only recently become a realistic possibility for mainstream organizations and that the underlying technology continues to develop at a lightning pace. Nevertheless, we hope to have shown that there is an urgent need both to a) continue to empirically investigate stakeholders' ideals around human-AI relational norms, as well as the causes and consequences of these, while also b) thinking explicitly about the risks and benefits of adopting (or failing to adopt) relationship-specific norms. Both of these agendas need to inform and constrain the other, along with practical and political considerations, toward a functional, implementable, and—as needed—revisable, reflective equilibrium.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

Implications for Organizational Research

The presented framework allows a novel look at how to study human and human-AI cooperation within organizations; see Reinecke et al. (in press) for a broader discussion of an empirical program using relationship functions to study human-AI interactions. One immediate next research step would be the integration of our framework in incentivized, game theoretical paradigms to study cooperation (Makovi et al., 2023). Cooperation would then not be measured via self-report but via actual behavior and real interactions between humans and AI. Such a line of research, ideally using representative samples of employees, would then also allow us to measure benchmarks of ideal relationship functions for human and AI cooperation, which could then be used as predictors for cooperation in other samples (Earp, McLoughlin, et al., 2021, for a similar approach). But one might also want to go beyond one-off interactions between two agents and start to investigate the development of relationship functions over time, when humans and AI repeatedly interact in the same social context. It would then also be important to examine whether such repeated human-AI interactions impact human cooperation. One optimistic prediction would be that while people hold reservations over cooperating with AI initially, repeated interactions will overcome fears and people will cooperate as well with AIs as humans. Thus, cooperation in human-AI collectives would not deteriorate over time. A bleaker prediction is based on recent work suggesting that interactions with AI might negatively impact subsequent interactions with other humans, thereby impeding human-human cooperation (Weisman & Janardhanan, 2020).

Acknowledgemet

Part of this work was carried out while Earp was based primarily at the Uehiro Centre for Practical Ethics, University of Oxford.

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG3-GV-2023-012).

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

References

Bankins, S, Ocampo, A C, Marrone, M, Restubog, S L D, & Woo, S E (2024). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*, 45(2), 159–182. https://doi.org/10.1002/job.2735
Google Scholar WorldCat

Bezrukova, K, Griffith, T L, Spell, C, Rice, V, & Yang, H E (2023). Artificial intelligence and groups: Effects of attitudes and discretion on collaboration. *Group & Organization Management*, 48(2), 629–670. https://doi.org/10.1177/10596011231160574
Google Scholar WorldCat

Brinkmann, L, Baumann, F, Bonnefon, J -F, Derex, M, Müller, T F, Nussberger, A -M, Czaplicka, A, Acerbi, A, Griffiths, T L, Henrich, J, Leibo, J Z, McElreath, R, Oudeyer, P -Y, Stray, J, & Rahwan, I (2023). Machine culture. *Nature Human Behaviour*, 7(11), Article 11. https://doi.org/10.1038/s41562-023-01742-2
Google Scholar WorldCat

Bryan, C J, Yeager, D S, & Hinojosa, C P (2019). A values-alignment intervention protects adolescents from the effects of food marketing. *Nature Human Behaviour*, 3(6), 596–603. https://doi.org/10.1038/s41562-019-0586-6
Google Scholar PubMed WorldCat

Bugental, D B (2000). Acquisition of the algorithms of social life: A domain-based approach. *Psychological Bulletin*, 126, 187–219. https://doi.org/10.1037/0033-2909.126.2.187
Google Scholar PubMed WorldCat

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

Choudhary, V, Marchetti, A, Shrestha, Y R, & Puranam, P (2023). Human-AI ensembles: When can they work? *Journal of Management*, 1–34. https://doi.org/10.1177/01492063231194968
Google Scholar WorldCat

Clark, M S, Lemay, E P, & Reis, H T (2020). Other people as situations: Relational context shapes psychological phenomena. In J F Rauthmann, R A Sherman, & D C Funder (Eds.), *The Oxford Handbook of Psychological Situations* (pp. 40–61). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.5
Google Scholar Google Preview WorldCat COPAC

Clark, M S, & Mills, J R (2012). A theory of communal (and exchange) relationships. *Handbook of Theories of Social Psychology*, 2, 232–250.
Google Scholar Google Preview WorldCat COPAC

Crandall, J W, Oudah, M, Tennom Ishowo-Oloko, F, Abdallah, S, Bonnefon, J -F, Cebrian, M, Shariff, A, Goodrich, M A, & Rahwan, I (2018). Cooperating with machines. *Nature Communications*, 9(1), Article 1. https://doi.org/10.1038/s41467-017-02597-8
Google Scholar WorldCat

Curry, O S, Mullins, D A, & Whitehouse, H (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, 60(1), 47–69. https://doi.org/10.1086/701478
Google Scholar WorldCat

Dafoe, A, Bachrach, Y, Hadfield, G, Horvitz, E, Larson, K, & Graepel, T (2021). Cooperative AI: Machines must learn to find common ground. *Nature*, 593(7857), 33–36. https://doi.org/10.1038/d41586-021-01170-0
Google Scholar PubMed WorldCat

Dietvorst, B J, Simmons, J P, & Massey, C (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033
Google Scholar PubMed WorldCat

Earp, B D, Lewis, J, Dranseika, V, & Hannikainen, I R (2021). Experimental philosophical bioethics and normative inference. *Theoretical Medicine and Bioethics*, 42(3–4), 91–111. https://doi.org/10.1007/s11017-021-09546-z
Google Scholar PubMed WorldCat

Earp, B D, McLoughlin, K L, Calcott, R, Caraccio, M, Monrad, J T, Owen, A M, Crockett, M J, & Clark, M S (2025). *How social relationships shape praise and blame: Strengthening and extending the relational norms model*.
https://www.researchgate.net/publication/387959265
Google Scholar Google Preview WorldCat COPAC

Earp, B D, McLoughlin, K L, Monrad, J T, Clark, M S, & Crockett, M J (2021). How social relationships shape moral wrongness judgments. *Nature Communications*, 12(1), Article 1. https://doi.org/10.1038/s41467-021-26067-4
https://doi.org/10.1038/s41467-021-26,067-4
Google Scholar WorldCat

Faul, F, Erdfelder, E, Buchner, A, & Lang, A -G (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149
Google Scholar PubMed WorldCat

Fehr, E, & Fischbacher, U (2003). The nature of human altruism. *Nature*, 425(6960), 785–791.
https://doi.org/10.1038/nature02043
Google Scholar PubMed WorldCat

FeldmanHall, O, Dalgleish, T, Thompson, R, Evans, D, Schweizer, S, & Mobbs, D (2012). Differential neural circuitry and self-interest in real vs hypothetical moral decisions. *Social Cognitive and Affective Neuroscience*, 7(7), 743–751.
https://doi.org/10.1093/scan/nss069
Google Scholar PubMed WorldCat

Fosgaard, T R, Hansen, L G, & Wengström, E (2023). Norm compliance in an uncertain world. *Journal of Behavioral and Experimental Economics*, 107, 102083. https://doi.org/10.1016/j.socec.2023.102083
Google Scholar WorldCat

Gabriel, I (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, 30(3), 411–437.
https://doi.org/10.1007/s11023-020-09539-2
Google Scholar WorldCat

Glikson, E, & Woolley, A W (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057
Google Scholar WorldCat

Haidt, J, & Baron, J (1996). Social roles and the moral judgment of acts and omissions. *European Journal of Social Psychology*, 26(2), 201–218. https://doi.org/10.1002/(SICI)1099-0992(199603)26:2<201::AID-EJSP745>3.0.CO;2-J
Google Scholar WorldCat

Hamilton, V L, & Sanders, J (1981). The effect of roles and deeds on responsibility judgments: The normative structure of wrongdoing. *Social Psychology Quarterly*, 44(3), 237–254. https://doi.org/10.2307/3033836
Google Scholar WorldCat

Harris-Watson, A M, Larson, L E, Lauharatanahirun, N, DeChurch, L A, & Contractor, N S (2023). Social perception in human-AI teams: Warmth and competence predict receptivity to AI teammates. *Computers in Human Behavior*, 145, 107765.
https://doi.org/10.1016/j.chb.2023.107765
Google Scholar WorldCat

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

Henrich, J, & Muthukrishna, M (2021). The origins and psychology of human cooperation. *Annual Review of Psychology*, 72(1), 207–240. https://doi.org/10.1146/annurev-psych-081920-042106
Google Scholar PubMed WorldCat

Kappes, A, Nussberger, A -M, Faber, N S, Kahane, G, Savulescu, J, & Crockett, M J (2018). Uncertainty about the impact of social decisions increases prosocial behaviour. *Nature Human Behaviour*, 2(8), 573–580. https://doi.org/10.1038/s41562-018-0372-x
Google Scholar PubMed WorldCat

Kellogg, K C, Valentine, M A, & Christin, A (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366–410. https://doi.org/10.5465/annals.2018.0174
Google Scholar WorldCat

Kolbjørnsrud, V, Amico, R, & Thomas, R J (2017). Partnering with AI: How organizations can win over skeptical managers. *Strategy and Leadership*, 45(1), 37–43. https://doi.org/10.1108/SL-12-2016-0085
Google Scholar WorldCat

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

Logg, J, Minson, J, & Moore, D A (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005
Google Scholar WorldCat

Mahmud, H, Islam, A K M N, Ahmed, S I, & Smolander, K (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, 175(121), 390. https://doi.org/10.1016/j.techfore.2021.121390
Google Scholar WorldCat

Makovi, K, Sargsyan, A, Li, W, Bonnefon, J -F, & Rahwan, T (2023). Trust within human-machine collectives depends on the perceived consensus about cooperative norms. *Nature Communications*, 14(1), Article 1. https://doi.org/10.1038/s41467-023-38592-5
Google Scholar WorldCat

March, C (2021). Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players. *Journal of Economic Psychology*, 87, 102426. https://doi.org/10.1016/j.joep.2021.102426
Google Scholar WorldCat

Mills, J, Clark, M S, Ford, T E, & Johnson, M (2004). Measurement of communal strength. *Personal Relationships*, 11(2), 213–230. https://doi.org/10.1111/j.1475-6811.2004.00079.x
Google Scholar WorldCat

Nielsen, Y A, Pfattheicher, S, & Keijsers, M (2022). Prosocial behavior toward machines. *Current Opinion in Psychology*, 43, 260–265.
Google Scholar PubMed WorldCat

Rahwan, I (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8
Google Scholar WorldCat

Rai, T S, & Fiske, A P (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, 118(1), 57–75. https://doi.org/10.1037/a0021867
Google Scholar PubMed WorldCat

Ranis, G, Stewart, F, & Ramirez, A (2000). Economic growth and human development. *World Development*, 28(2), 197–219.

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025

https://doi.org/10.1016/S0305-750X(99)00131-X
Google Scholar WorldCat

Reinecke, M G, Kappes, A, Porsdam Mann, S, Earp, B D, & Savulescu, J (in press). *T*he need for an empirical research program regarding human-AI relational norms. *AI and Ethics*.
Google Scholar WorldCat

Savulescu, J, Gyngell, C, & Kahane, G (2021). Collective Reflective Equilibrium in Practice (CREP) and controversial novel technologies. *Bioethics*, 35(7), 652–663. https://doi.org/10.1111/bioe.12869
Google Scholar PubMed WorldCat

Smith, N C (2003). Corporate social responsibility: Whether or how? *California Management Review*, 45(4), 52–76. https://doi.org/10.2307/41166188
Google Scholar WorldCat

Tomasello, M (2020). The moral psychology of obligation. *Behavioral and Brain Sciences*, 43, e56. https://doi.org/10.1017/S0140525X19001742
Google Scholar WorldCat

Vanneste, B S, & Puranam, P (2024). Artificial intelligence, trust, and perceptions of agency. *Academy of Management Review*, 1–19. https://doi.org/10.5465/amr.2022.0041
Google Scholar WorldCat

Wagner, C, Strohmaier, M, Olteanu, A, Kıcıman, E, Contractor, N, & Eliassi-Rad, T (2021). Measuring algorithmically infused societies. *Nature*, 595(7866), Article 7866. https://doi.org/10.1038/s41586-021-03666-1
Google Scholar WorldCat

Wang, W, Gao, G (Gordon), & Agarwal, R (2023). Friend or foe? Teaming between artificial intelligence and workers with variation in experience. *Management Science*, 70(9), 5753–5775. https://doi.org/10.1287/mnsc.2021.00588
Google Scholar WorldCat

Weisman, H, & Janardhanan, N S (2020). The instantaneity of gendered voice assistant technology and manager perceptions of subordinate help. *Academy of Management Proceedings*, 2020(1), 21149. https://doi.org/10.5465/AMBPP.2020.21149abstract
Google Scholar WorldCat

© Oxford University Press 2025

Downloaded from https://academic.oup.com/edited-volume/59762/chapter/508608206 by Andreas Kappes on 04 April 2025