Three-Dimensional Motivation Theory of Algorithm Rejection

Authors: Zhang Yuyan, Xu Liying, Yu Feng, Ding Xiaojun, Wu Jiahua, Zhao Liang, Xu Liying, Yu Feng

Date: 2021-11-16T00:00:00+00:00

Abstract

Algorithm aversion refers to the phenomenon wherein, although algorithms can typically make more accurate decisions than humans, people still prefer human decision-making. The three-dimensional motivational theory of algorithm aversion identifies three main causes: suspicion of algorithmic agents, lack of moral status, and annihilation of human characteristics, which correspond respectively to three psychological motivations—trust, responsibility, and control—and align with three feasible approaches to mitigate algorithm aversion: enhancing human trust in algorithms, strengthening algorithmic agency responsibility, and exploring personalized algorithmic design to highlight human control over algorithmic decision-making. Future research could further investigate the boundary conditions of algorithm aversion and other potential motivations from a more social perspective.

Full Text

A Three-Dimensional Motivational Theory of Algorithm Aversion

ZHANG Yuyan[1], XU Liying[2], YU Feng[1], DING Xiaojun[3], WU Jiahua[2], ZHAO Liang[4]

[1] Department of Psychology, School of Philosophy, Wuhan University, Wuhan 430072, China
[2] Department of Psychology, School of Social Sciences, Tsinghua University, Beijing 100084, China
[3] Department of Philosophy, School of Humanities and Social Science, Xi' an Jiaotong University, Xi' an 710049, China
[4] Department of Publishing Science, School of Information Management, Wuhan University, Wuhan 430072, China

Abstract: Algorithm aversion refers to the phenomenon where people prefer human decisions despite algorithms typically making more accurate decisions. The three-dimensional motivational model of algorithm aversion identifies three primary causes: doubt regarding algorithmic agency, lack of moral standing, and the annihilation of human uniqueness. These correspond respectively to three psychological motivations—trust, responsibility, and control—and suggest three feasible approaches to mitigate algorithm aversion: enhancing human trust in algorithms, strengthening algorithmic agency responsibility, and exploring personalized algorithm design to highlight human control over algorithmic decisions. Future research should further investigate the boundary conditions of algorithm aversion and other potential motivations from a more social perspective.

Keywords: algorithmic decision-making, algorithm aversion, psychological motivation, human-robot interaction

Classification: B849:C91

1 Introduction

Decision-making is ubiquitous in daily life. Traditionally, humans have been the primary agents of decision-making, whether as principals or advisors. However, human decision-making capabilities are limited by factors such as experience and emotion, often leading to errors and undesirable outcomes. With the development and proliferation of algorithmic technology, algorithmic decision-making has been widely introduced into professional and everyday life to overcome these limitations. Algorithmic decision-making serves as an umbrella term encompassing augmented decision-making, decision aids, decision support systems, expert systems, decision formulas, and computer assistance (Burton et al., 2020), with recent research extending the concept to machine decision-making, robot decision-making, and artificial intelligence decision-making (Malle et al., 2015). Compared to human decision-making, algorithmic decision-making offers numerous advantages: speed (Bonnefon et al., 2016), accuracy (Donnelly, 2017; Lohr, 2016), objectivity (Andrews et al., 2006), universality (Esteva et al., 2017), and low cost (Common Cents Lab, 2017). These characteristics have led to increasing adoption across diverse societal domains, including healthcare (Biró et al., 2021; van den Berg et al., 2017), economics (Harvey et al., 2017; Kaya et al., 2017), judiciary (Angwin et al., 2016), transportation (Badue et al., 2020; Fournier, 2016), military (Horowitz, 2016; Shin et al., 2019), and daily life (Roberts, 2017).

While algorithmic technology has long liberated humans from physically demanding tasks (Parasuraman & Riley, 1997), algorithmic decision-making is now increasingly performing cognitive functions such as judgment, planning, and creative thinking (Chouard, 2016; Luo et al., 2019; Oliveira, 2017). Indeed, advances in algorithmic technology have brought revolutionary changes to society.

Nevertheless, these advantages have not endeared algorithmic decision-making to people. Philosopher Bostrom (2014) warned that machines making decisions on humanity' s behalf could lead to catastrophe, while Elon Musk described the rise of autonomous machines as humanity' s "biggest existential threat" (McFarland, 2014). The general public similarly prefers human over algorithmic decision-making. Research demonstrates that although algorithms typically outperform humans in decision-making tasks, people still choose human decisions or follow human advice. Dietvorst et al. (2015) termed this phenomenon—where people prefer human decisions while avoiding more accurate algorithmic decisions—algorithm aversion. Algorithm aversion does not only emerge after people learn that algorithmic decisions may err; it more frequently occurs before they even understand algorithm performance. Thus, algorithm aversion essentially represents a biased evaluation of algorithms, manifesting as negative attitudes and behaviors toward them (Jussupow et al., 2020).

Algorithm aversion manifests across cognitive, emotional, and behavioral dimensions. Cognitively, it appears as distrust of algorithmic capabilities (Prahl & Van Swol, 2017). For instance, despite evidence that algorithmic decision-making in judicial systems could counteract inherent human subjectivity and potential errors (Andrew et al., 2009), 56% of respondents in a Pew survey still considered algorithmic risk assessment for parole decisions unacceptable (Smith, 2018). In healthcare, patients more harshly judge physicians who seek advice from algorithms rather than colleagues (Shaffer et al., 2013). Emotionally, algorithm aversion appears as negative feelings toward using algorithmic decisions or their outcomes (Lee, 2018; Leyer & Schneider, 2019) and even moral condemnation (Voiklis et al., 2016). When evaluating identical artworks, people prefer those purportedly created by humans over algorithms (Jago, 2019). When errors occur in judicial decisions, people exhibit stronger negative emotional reactions to algorithmic errors and are more likely to take extralegal measures to disrupt judicial order when courts fail (Ireland, 2020). Behaviorally, people choose algorithmic decisions less frequently or utilize them to a lesser degree compared to human decisions, sometimes even resisting algorithmic outcomes (Filiz et al., 2021). In stock market experiments, people trust identical predictions more when attributed to human experts and are more willing to accept human advice (Önkal et al., 2009). In healthcare, consumers are less willing to use AI-provided medical services compared to human-provided services and are willing to pay less for them (Longoni et al., 2019).

Why do humans reject algorithms across these psychological dimensions? We propose that when facing this novel decision-making form, humans intuitively ask three sequential questions. First, can algorithms truly make decisions and do they have the capacity to do so? The answer is typically negative, as people generally doubt algorithmic capabilities, generating suspicion and distrust that leads to algorithm aversion. Second, even if algorithms can make decisions, do we really need them to, or what benefit does algorithmic decision-making offer individuals? Again, the answer is usually negative, because humans tend to shift responsibility when making decisions, yet algorithms lack moral agency and the capacity to bear responsibility, making algorithmic decision-making seem useless and thus prompting rejection. Third, even if algorithms are capable and can share moral responsibility, does this decision-making model truly benefit humanity? The answer remains negative, as algorithmic decision-making can cause people to lose control, leading to feelings of personal characteristic annihilation and dehumanization, ultimately resulting in algorithm rejection. We therefore propose a three-dimensional motivational theory of algorithm aversion to explain its causes and explore potential remedies (see Table 1).

2 Causes of Algorithm Aversion

Algorithmic decision-making is fundamentally a human-computer interaction process, and the causes of algorithm aversion inevitably involve humans, algorithms, and their interaction. At a deeper level, we identify three core causes: algorithm agent doubt, moral standing deficiency, and human characteristic annihilation. These manifest as tendencies to doubt algorithmic capabilities as a novel entity, to view algorithms as unable to bear moral responsibility, and to perceive algorithmic generalization as diminishing human uniqueness. These address the questions of "whether they can" (not understanding or believing in algorithmic decision-making capacity), "whether to use them" (uselessness due to unassignable responsibility), and "whether they are good" (dissolving individual uniqueness), corresponding to trust/distrust, responsibility-shirking/acceptance, and control/loss-of-control motivations.

2.1 Algorithm Agent Doubt

The first question concerns whether algorithmic decision-making can be effective, to which people tend to respond negatively, expressing doubt about algorithms. The first cause of this doubt may be unfamiliarity with algorithms. Few technologies are immediately accepted upon workplace introduction (Parasuraman & Riley, 1997). Initial unfamiliarity with new technology often breeds dislike or distrust of automated systems. However, as algorithmic decision-making becomes more prevalent and people gain exposure, acceptance increases. This phenomenon can be explained by the mere exposure effect: repeated exposure to external stimuli such as images and sounds increases liking (Zajonc, 1968). For example, most people accept weather forecasts from meteorological models rather than neighbors because such models have been widely used for decades. Conversely, algorithmic fashion recommendations remain relatively novel and face greater resistance (Logg et al., 2019). Japanese individuals, more familiar with robots in daily life, more readily accept machine moral decision-making (Komatsu, 2016). Ireland (2020) suggests that as people become accustomed to algorithms, algorithm aversion in judicial systems may weaken or disappear entirely, with algorithmic errors viewed no differently than human judicial errors.

Two factors influence familiarity with algorithmic decision-making: algorithmic transparency and human expertise. First, algorithmic opacity may hinder familiarity. Algorithmic decisions emerge from black-box methods (Castelvecchi, 2016), with internal workings unknown to users (Kroll et al., 2017; Lin & Tang, 2020). The reasoning behind statistical predictions may be inaccessible or mysterious to those without statistical training (Önkal et al., 2009). Research finds that individuals with weaker mathematical abilities show lower acceptance of algorithmic decisions (Logg et al., 2019). Interestingly, both algorithmic and human decision-making are opaque, yet people believe they can understand human decision-making processes through introspection. This coupling effect reduces perceived opacity for human decisions while amplifying it for algorithmic decisions. This asymmetry in transparency perception undermines confidence in understanding algorithmic decisions, causing people to underestimate their comprehension of algorithms while overestimating their understanding of human decisions, thereby strengthening algorithm aversion (Cadario et al., 2021). Explanations of algorithmic processes can improve both objective understanding and subjective confidence, thereby increasing algorithmic decision utilization (Cadario et al., 2021; Yeomans et al., 2019).

Second, domain familiarity and expertise affect acceptance. Financially educated individuals process financial information more effectively and thus more readily accept algorithmic advice (Lusardi & Mitchell, 2011), whereas less educated individuals less recognize their need for advice (Lee & Moray, 1992). Those with less investment knowledge more strongly reject algorithmic investment assistance (Niszczota & Kaszás, 2020). However, overconfident individuals may also reject algorithmic decisions (Soll & Mannes, 2011). When Meehl (1954) first proposed that algorithms outperform humans in certain decisions, experts vehemently objected, unwilling to believe linear models could surpass their judgment and even responding with hostility. In prediction tasks, experts fail to recognize algorithmic advice value, ignoring provided information and stubbornly maintaining initial judgments. Yet experts who reject algorithmic advice perform worse than laypeople (Logg et al., 2019). These findings suggest a non-linear relationship between domain familiarity/expertise and algorithm aversion, where both weak and excessive familiarity/expertise strengthen rejection. Future research should examine this relationship across varying expertise levels.

The second cause of algorithm agent doubt may be perceived lack of decision-making competence. People generally doubt algorithmic expertise, believing algorithmic performance inferior to human decision-making (Dzindolet et al., 2002; Prahl & Van Swol, 2017). For instance, people resist medical algorithmic advice because they doubt algorithms possess adequate medical expertise (Promberger & Baron, 2006). Even when evaluating identical artworks, people prefer those purportedly created by humans (Jago, 2019), revealing a bias that algorithmic expertise is inferior. People prefer human decisions even when algorithmic and human accuracy are comparable or when algorithms make equivalent errors (Gogoll & Uhl, 2018), only choosing algorithms when they demonstrate clear superiority (Bigman & Gray, 2018). Sometimes people reject algorithms despite knowing their superior performance (Grove & Lloyd, 2006). While expertise gaps could be bridged through learning, people tend to believe algorithms lack learning capacity (Highhouse, 2008). Specifically, people view algorithmic errors as systematic and unlearnable, whereas human errors are random and improvable over time. Consequently, when algorithms err, people are more likely to bypass them entirely (Filiz et al., 2021). This perception may be strengthened by algorithmic opacity, where people cannot trace decision processes (Angwin et al., 2016; O' Neil, 2017).

2.2 Moral Standing Deficiency

The second cause of algorithm aversion is algorithms' lack of moral standing, preventing them from bearing post-decision responsibility. Often, people follow others' decisions to transfer or dilute their own responsibility. If algorithms lack moral agency, they become useless for responsibility transfer, causing algorithm aversion. This effect is particularly pronounced in moral versus non-moral domains. For life-and-death moral decisions, even when machines produce positive outcomes, people prefer average humans over superior machines (Bigman et al., 2019). In economic games, people prefer delegating decisions affecting others' payoffs to humans rather than machines, believing such tasks involve morality and instinctively disliking machine involvement in moral domains (Gogoll & Uhl, 2018). In investment, people prefer human fund managers, especially for morally charged investments, showing stronger algorithm aversion because they believe moral competence is required to judge controversial investments—capacity they attribute to humans but not machines (Niszczota & Kaszás, 2020). This may stem from moral decisions being deeply rooted in emotion and requiring full moral standing (Gray et al., 2017; Haidt, 2001). People judge moral standing based on perceived mind differences (Bastian et al., 2012; Gray et al., 2012).

Mind is perceived along two dimensions: agency and experience (Gray et al., 2007). Agency involves thinking, reasoning, planning, and implementing intentions, while experience involves feeling emotions and sensations like pain (Gray et al., 2012).

Algorithms are perceived as possessing some moral agency, but far weaker than humans. Algorithms exhibit some agency (Gray & Wegner, 2012) through complex calculations, yet full agency requires self-control, planning, communication, and thought (Gray et al., 2007). Other research suggests moral decision-making agents require interactivity, autonomy, and adaptability (Floridi & Sanders, 2004), plus moral reasoning, autonomous action, communication, and consequence judgment (Cushman, 2008; Malle, 2016; Malle & Scheutz, 2014). Machines lack these capacities, leading people to deny their moral decision-making ability. Beyond agency, experience is crucial for moral decisions. Emotions are vital for moral judgment (Greene et al., 2001; Haidt, 2001; Haidt et al., 1993; Koenigs et al., 2007), particularly empathy—the capacity to feel others' pain—which seems central to moral judgment (Aaltola, 2014; Decety & Cowell, 2014). Machines appear to lack genuine emotional capacity (Reinecke et al., 2021). Thus, despite some agency, machines lack experience and the ability to feel moral emotions (Malle & Scheutz, 2014), rendering them incapable of moral decision-making.

Most importantly, insufficient moral agency prevents algorithms from bearing moral responsibility. If algorithms cannot assume moral responsibility, humans lack the option to transfer responsibility to them, making algorithmic decision-making unattractive from a responsibility-sharing perspective (Bonaccio & Dalal, 2006). When following human advisors, people feel responsibility shifts to the advisor, but this transfer does not occur with algorithms because algorithms are perceived as lacking responsibility-bearing capacity (Bonaccio & Dalal, 2006). Armstrong (1980) found that people trust experts despite evidence that expert judgment is sometimes no more accurate than lay judgment, a phenomenon observed across political forecasting, conflict outcome prediction, and stock market forecasting (Green & Armstrong, 2007; Tetlock, 2009). Armstrong (1980) attributed this to responsibility transfer: if predictions prove inaccurate, experts bear more blame, allowing decision-makers to shift responsibility. Bonaccio and Dalal (2006) argued that such responsibility-shifting motives only operate when advisors are human, not algorithmic, because humans are seen as capable of bearing responsibility while algorithms are not (Promberger & Baron, 2006). In healthcare, when patients must decide on important medical procedures, they transfer responsibility to human doctors because following medical advice reduces their sense of responsibility, whereas following algorithmic advice does not (Promberger & Baron, 2006).

2.3 Human Characteristics Annihilation

The third cause of algorithm aversion is the perceived symbolic threat to human identity or uniqueness, and the associated loss of control. Individually, people typically view themselves as unique, a tendency especially pronounced in individualistic cultures (Brewer, 1991). Algorithms, being relatively rational, operate in standardized and patterned ways without personal data, processing every case identically (Haslam, 2006). This mismatch between basic beliefs creates resistance. Compared to human healthcare providers, consumers are less willing to use AI medical services because AI raises concerns that individuals' unique characteristics, circumstances, and symptoms will be ignored—a concern termed uniqueness neglect (Longoni et al., 2019). In consumer contexts, information asymmetry between product features and consumer needs damages firm-consumer relationships (Van Swol, 2009). People are less willing to use algorithms for subjective tasks, preferring friends over algorithms for movie, book, or joke recommendations because these decisions are governed by personal taste (Yeomans et al., 2019). Conversely, in domains with clear external accuracy standards like sports forecasting, people may trust algorithmic advice more (Logg et al., 2019).

The deeper cause of personality loss may be concern about annihilation of human group identity and uniqueness—perceiving algorithmic decision-making, particularly highly autonomous algorithms, as threatening human in-group distinctiveness and values, even feeling dehumanized. Humans consider themselves distinct from other groups, yet highly autonomous algorithmic control may blur boundaries between "human" and "tool." People dislike robots that autonomously judge and decide without human instruction, perceiving greater threat (Złotowski et al., 2017). Graduate admissions based solely on data rather than face-to-face conversation are considered inhumane (Dawes, 1979). Increasing human participation may enhance algorithm acceptance. People prefer combined algorithm-expert decisions over algorithm-only decisions, provided algorithm assistance does not replace human judgment. In fact, people may prefer experts using algorithmic assistance to those without it (Palmeira & Spassova, 2015). Pezzo et al. (2006) found that physicians using algorithmic support receive less negative evaluation when erring than those without support. Even imperfect expert involvement may facilitate better decision-making (Kuncel, 2008). Research shows that when algorithms assist rather than replace human decisions, and humans retain final authority, algorithm aversion weakens. People more readily accept and like consulting algorithms than being subject to algorithmic decisions (Dietvorst et al., 2015). Restricting machines to subordinate roles with humans making final decisions reduces algorithm aversion (Bigman & Gray, 2018). Longoni et al. (2019) similarly found no resistance to medical AI when it assisted rather than replaced physicians. When people can autonomously modify algorithmic outcomes to retain final decision authority, they more readily accept algorithmic decisions (Dietvorst et al., 2018).

3 Changing Algorithm Aversion

Algorithms and algorithmic decision-making inevitably permeate human life in the AI era, with many algorithms already driving information acquisition and understanding (e.g., website recommendation algorithms) without users' awareness. Algorithmic decision-making will undoubtedly gain increasing future application. While still rapidly developing and immature in some domains, algorithmic decision-making already surpasses human capability in many areas, yet humans have not accepted it for various reasons. If people could appreciate algorithmic utility and increase acceptance, it could bring more effective solutions to practical problems and generate social economic benefits. Addressing the three causes of algorithm aversion—algorithm agent doubt, moral standing deficiency, and human characteristic annihilation—which correspond to the three psychological motivations of trust, responsibility, and control, we propose three approaches to increase algorithm acceptance: enhancing human trust in algorithms, strengthening algorithmic agency responsibility, and exploring personalized algorithm design to highlight human control.

3.1 Enhancing Trust in Algorithms

The core factor countering algorithm agent doubt is likely trust. Trust permeates most causes of algorithm aversion. Human cognitive characteristics—including insufficient algorithmic experience, inadequate perceived algorithmic capability, and insufficient involvement in algorithmic decision-making—can lead to distrust or fragile trust, prompting algorithm aversion. Distrust manifests in two ways. First, initial trust in algorithmic decision-making is lower than in human decision-making. Even before learning about algorithmic imperfections or experiencing errors, people trust algorithms less (Longoni et al., 2019). Second, trust in algorithms is more fragile, meaning one or two negative experiences damage trust more severely than with humans (Prahl & Van Swol, 2017). Consequently, people tolerate human errors but reject algorithmic ones. Dietvorst et al. (2018) found that people have lower tolerance for algorithmic errors, with confidence plummeting after observing them and subsequent choices avoiding algorithms. Thus, much algorithm aversion stems from failure to establish trust in algorithmic decision-making (Lee & Seppelt, 2006)—algorithms have not "convinced" humans of their decision-making capacity.

For algorithm aversion research, the simplest trust-building approach is enhancing human algorithmic expertise and familiarity while demonstrating algorithmic capability. As algorithmic and AI elements increasingly appear in daily life, human knowledge and familiarity will gradually increase until algorithms become commonplace. At that point, demonstrating capability becomes crucial. Professional and effective algorithmic advice increases trust and utilization (Goodyear, 2016; Kramer et al., 2018). Revolutionary transformation of healthcare through AI medical services requires expert-level accuracy (Leachman & Merlino, 2017). Furthermore, when people can autonomously adjust algorithmic outputs according to their needs (Greene et al., 2016), their perception of algorithmic learning and adaptation capacity strengthens. Although algorithmic decision-making remains imperfect, people are willing to accept error-prone algorithms that can learn and adjust (Berger et al., 2021). Currently, however, algorithmic expertise seems only salient when clearly superior to human performance (Bigman & Gray, 2018).

3.2 Strengthening Algorithmic Agency Responsibility

People may prefer human decisions over algorithmic ones due to responsibility-sharing and -shirking motives. Such motives require algorithms to bear moral responsibility, necessitating strengthened moral standing—making algorithms more human-like. This "human-likeness" can be achieved in two ways: making them appear human or making them seem to possess defining human capacities. The latter involves enhancing perceived algorithmic mental capacities to increase acceptance. When algorithms or AI appear more human-like in mental capacity, people more strongly believe they can fully perform intended functions (Waytz et al., 2014). Conscious agents are perceived as better able to control their behavior, successfully complete tasks through conscious prediction and planning, and take responsibility for outcomes (Cushman, 2008). For example, increasing perceived emotional similarity effectively increases algorithm use in subjective tasks (Castelo et al., 2019).

For algorithmic decision-making embodied in AI agents like machines, robots, and autonomous vehicles, anthropomorphism can enhance human-likeness and moral agency. Anthropomorphism is the psychological process of attributing uniquely human characteristics to non-human objects (Waytz et al., 2010; Xu et al., 2017), achievable through physical features (face, eyes, body, movement), psychological features (preferences, humor, personality, emotions, empathy), language (speech, language recognition), social dynamics (cooperation, encouragement, question-answering, reciprocity), and social roles (doctor, teammate, opponent, teacher, pet, guide) (Fogg, 2002). Importantly, anthropomorphism influences mind perception, particularly agency and experience (Gray et al., 2017). Research shows that anthropomorphizing AI agents in mind perception increases trust. For autonomous vehicles, when they appear to perceive and think about their environment rather than being mindless machines, they seem more capable of navigating traffic (Waytz et al., 2014). Anthropomorphizing external features can indirectly enhance perceived mental capacity and trust. When autonomous vehicles are given names, genders, and human-like voices, people trust them more, blame them less for accidents caused by others, and attribute success more to them when they avoid accidents (Waytz et al., 2014).

However, excessive anthropomorphism or mental capacity perception may backfire. When algorithms or AI agents become too human-like in appearance or mental capacity, affinity drops sharply, producing discomfort and fear (Gray & Wegner, 2012)—the uncanny valley effect (Mori, 1970)—leading to algorithm aversion. Therefore, when attributing uniquely human characteristics to algorithms, these should be limited to a comfortable range to increase acceptance and trust.

It must be noted that although research demonstrates algorithmic superiority over humans in prediction tasks (Jordan & Mitchell, 2015), chess games (Dockrill, 2017; Silver, 2017), and stock market return maximization (Zuckerman, 2019), morality is relatively subjective (Yu & Han, 2018). No clear evidence shows whether algorithms can make better moral decisions, as humans hold diverse positions (deontological vs. utilitarian) on moral dilemmas. Algorithm acceptance means appreciating algorithmic utility and holding positive attitudes and behaviors (Jussupow et al., 2020), not blind acceptance—particularly regarding moral issues, which warrant cautious consideration. Even if anthropomorphism reduces algorithm aversion by increasing perceived responsibility-bearing capacity, it may not fully alleviate responsibility attribution concerns. The more important question concerns responsibility attribution for algorithmic errors: who should be responsible—designers, manufacturers, or users? As algorithmic decision-making proliferates, this inevitable question represents the most effective way to reduce responsibility attribution concerns. Even if algorithms can share responsibility, how to punish them remains challenging. Parallel solutions may involve moral education for humans to encourage greater responsibility acceptance and less shirking.

3.3 Exploring Personalized Algorithm Design

Human characteristic annihilation manifests as inability to demonstrate uniqueness, with the underlying motivation being loss of control over the external world or perceived autonomy loss from algorithmic decision-making. To restore human control and sense of unique existence, exploring personalized algorithm design to highlight human control may be optimal. This can manifest through satisfying personalized needs or granting humans final decision authority.

First, algorithmic decision-making becomes more acceptable when algorithms understand personal preferences. In consumer contexts, electronic product selection tools that predict consumer-perceived product attractiveness can recommend preferred products from vast catalogs (Diehl et al., 2003), reducing search time while increasing sales likelihood (Senecal & Nantel, 2004). In healthcare, attention to uniqueness—where medical AI provides more personalized and targeted services—reduces resistance (Longoni et al., 2019). In investment, more personalized advice increases consumer preference and trust (Alserda et al., 2019; Lourenço et al., 2020).

While personalized algorithmic decision-making is already trending (e.g., news recommendation algorithms), it creates new problems. First is the echo chamber effect (Cinelli et al., 2021). Current information media extensively employ algorithms based on the principle that preference for certain content increases its recommendation frequency. These preferences may derive from historical co-occurrence or simple tags. Regardless, this echo chamber effect weaves an information cocoon that shields other information, potentially leading to intellectual stagnation. Second is privacy concerns. Personalized algorithmic decision-making requires personal data collection, yet people do not want personalized services in all domains. Research shows positive attitudes toward personalization in commercial applications (shopping, entertainment) but opposition in news sources, social media, political campaigns, and regarding sensitive personal information (Ipsos Mori, 2020; Pew Research Center, 2019). People need not just personalized algorithms but transparent ones that respect data privacy and allow user adjustment (Kozyreva et al., 2021). Over-collecting data to enhance personalization, violating privacy, or overstepping boundaries in sensitive domains may prove counterproductive.

Second, humans should retain final authority in decision-making systems. As machines increasingly share human environments, complete algorithmic replacement of human decision-making or jobs is never optimal. Humans and algorithms excel at different aspects: algorithms may "do things right" better than humans, but only humans know "what to do" and "how to do it." Leveraging algorithmic strengths in computation and information integration to compensate for human weaknesses, while applying human wisdom where algorithms fall short, creates synergy. Algorithm-assisted or human-algorithm collaborative decision-making, ensuring human primacy, fosters more open attitudes toward algorithmic decision-making. Thus, strengthening human-algorithm collaborative decision-making while ensuring human agency makes people more receptive. Compared to either algorithm-only or human-only decisions, people may believe human-algorithm collaboration yields superior decisions (Palmeira & Spassova, 2015; M. V. Pezzo & S. P. Pezzo, 2006), provided humans retain final authority (Starke & Lünich, 2020). Chinese scholar Qian Xuesen proposed the Hall for Workshop of Metasynthetic Engineering decades ago, advocating human-centered, human-computer integration where computers assist rather than replace humans, combining computers' high-speed information processing with humans' comprehensive thinking (including logical, imaginative, and creative thinking) (Huang, 2005). Qian' s human-centric approach, prioritizing human wisdom and interests (Li & Jiang, 2019), remains profoundly relevant for contemporary algorithmic decision-making research.

4 Discussion and Conclusion

The three-dimensional motivational theory of algorithm aversion simulates humans' intuitive thinking framework when facing algorithmic decisions—the sequential questions humans intuitively ask: whether we understand and believe algorithms can decide, whether using them leaves us bearing responsibility for failures, and whether using them diminishes our uniqueness as humans. These three questions reveal the motivations of trust/distrust, responsibility acceptance/shirking, and control/loss-of-control. While we have reviewed existing research and attempted theoretical construction, much remains to be explored.

First, this theory may implicitly suggest that algorithmic decision-making is superior to human decision-making and that accepting algorithms is "normative." In reality, humans do not decide purely rationally, and what makes us human may be precisely our unique rationality beyond machines. Research finds that in the AI era, humans devalue capacities where algorithms excel (general cognitive ability, physical ability, negative emotions) while emphasizing capacities like morality and aesthetics to distinguish themselves from machines (Yu, 2020). Human brilliance sometimes shines through emotion rather than machine rationality. Moreover, human acceptance or rejection of something does not necessarily stem from comparative advantage. Humans do not accept systems merely because they are superior, nor abandon products simply because others are more cost-effective. Avoiding normative judgments based purely on descriptive superiority is itself a human characteristic; otherwise, humans would be no different from algorithms. This theory does not presuppose such a normative stance but rather investigates why humans reject algorithms despite their superiority in some domains. While the theory does suggest leveraging algorithmic advantages to optimize human decision-making, it does not advocate replacing human decision-making with algorithms.

Second, the three-dimensional motivational theory is open to other possible motivations. Broader cognitive, existential, and social motivations may serve as alternatives (Jost et al., 2003; Jost et al., 2009). While trust, responsibility, and control belong to generalized social or cognitive motivations, other possibilities exist. For example, need for cognitive closure involves desire for clear, perfect answers, causing high-need-for-closure individuals to feel uncomfortable with openness and uncertainty, thus disliking unresolved decisions (Otto et al., 2016). Algorithmic decision-making, as a product of the modern information era, may represent openness and uncertainty to such individuals, triggering algorithm aversion. Research indeed finds algorithm aversion may emerge after seeing algorithms err (Dietvorst et al., 2020), possibly due to conflict between need for perfect answers and the disappointment of algorithmic failure. Another example is social identity. Traditional intergroup theory identifies human identity as the broadest social identity, reflecting self-categorization as human and differentiation from non-humans (Turner et al., 1987). Recent research shows social identity expanding to non-human entities, such as psychological connections with animals comprising solidarity, pride, and perceived similarity (Amiot et al., 2020). Such connections clearly generate positive emotional bonds and preference for animals. Analogously, lack of psychological connection with algorithms (like with animals) may cause dislike or even aversion. For instance, people may reject AI doctors due to weaker emotional connection compared to human doctors. Psychological connection also matters for understanding decision processes. Although both human and algorithmic decision-making are opaque, perceived similarity through psychological connection leads people to believe they can understand human decisions through introspection (Nisbett & Wilson, 1977), however inaccurate this belief may be (Kahneman, 2003; Morewedge & Kahneman, 2010). This belief preferences human decisions, while lack of connection prevents understanding algorithmic processes and prompts rejection. This requires further research and theoretical refinement.

Third, emphasizing psychological motivations does not place them in absolute dominance among all factors influencing algorithm aversion. Human rejection of algorithmic decisions may also stem from various subjective reasons or practical considerations including legal, political, social context, and philosophical reflection. Legal factors include "data ethics" or "AI ethics" frameworks, privacy boundaries, individual rights and freedoms, and technological oligopolies (Peng, 2020). Social factors include concerns about unfair resource distribution, technological unemployment, class solidification, and social exclusion (Li, 2021; Sun, 2021). Philosophers critically reflect on algorithm aversion, arguing that technological expansion may create existential deficits, erode meaning and value, and render existence absurd (Zhou, 2021). While the three-dimensional motivational theory emphasizes psychological factors, it does not deny other important influences. Algorithm aversion is necessarily a multifaceted phenomenon, but this theory specifically examines its causes and solutions from psychological motivational perspectives.

In summary, algorithm aversion may stem from three-dimensional causes—algorithm agent doubt, moral standing deficiency, and human characteristic annihilation—reflecting motivations of trust, responsibility, and control. Future research should explore boundary conditions of algorithm aversion and identify psychological variables that change or even reverse the phenomenon. Additionally, further investigation of other possible psychological motivations, empirical testing, and theoretical integration is needed. Finally, as algorithmic decision-making increasingly manifests through AI agents, transitioning from instrumental to social entities, examining attitudes and corresponding psychological motivations from a social agent perspective is advisable.

References

Huang, Z. (2005). Human-centered, human-computer integrated, from qualitative to quantitative metasynthetic method. *Journal of Xi' an Jiaotong University (Social Sciences Edition)*, 25(2).

Li, M. (2021). *Automating inequality: How high-tech tools profile, police, and punish the poor* (pp. 148-171). Beijing: The Commercial Press.

Li, Y., & Jiang, X. (2019). Qian Xuesen and the AI boom of the 1980s. *Journal of Xi' an Jiaotong University (Social Sciences Edition)*, 39(6).

Peng, X. (Ed.). (2020). *Taming algorithms: Digital discrimination and algorithmic regulation* (pp. 35-41). Shanghai: Shanghai People' s Publishing House.

Sun, W. (2021). The "labor" of intelligent systems and its social consequences. *Philosophical Research*, (8), 30-40+128.

Xu, L., Yu, F., Wu, J., Han, T., & Zhao, L. (2017). Anthropomorphism: From "it" to "he" . *Advances in Psychological Science*, 25(11), 1942-1954.

Yu, F. (2020). On artificial intelligence and being human. *People' s Tribune · Academic Frontier*, (1), 30-36.

Yu, F., & Han, T. (2018). A model of limited moral objectivism. *Journal of Tsinghua University (Philosophy and Social Sciences Edition)*, 33(157), 148-163+193.

Zhou, L. (2021). The philosophical nature and critical transcendence of intelligent fetishism. *Philosophical Research*, (8), 41-50.

Aaltola, E. (2014). Affective empathy as core moral agency: Psychopathy, autism and reason revisited. *Philosophical Explorations*, 17(1), 76-92.

Alserda, G. A., Dellaert, B. G., Swinkels, L., & van der Lecq, F. S. (2019). Individual pension risk preference elicitation and collective asset allocation with heterogeneity. *Journal of Banking & Finance*, 101, 206-225.

Amiot, C. E., Sukhanova, K., & Bastian, B. (2020). Social identification with animals: Unpacking our psychological connection with other animals. *Journal of Personality and Social Psychology*, 118(5), 991-1017.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7-27.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. Retrieved May 19, 2021, from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Armstrong, J. S. (1980). The seer-sucker theory: The value of experts in forecasting. *Technology Review*, 82(7), 16–24.

Badue, C., Guidolini, R., Carneiro, R. V., Azevedo, P., Cardoso, V. B., Forechi, A., ···& De Souza, A. F. (2020). Self-driving cars: A survey. *Expert Systems with Applications*, 113816.

Bastian, B., Loughnan, S., Haslam, N., & Radke, H. R. (2012). Don' t mind meat? The denial of mind to animals used for human consumption. *Personality and Social Psychology Bulletin*, 38(2), 247–256.

Berger, B., Adam, M., Rühr, A., & Benlian, A. (2021). Watch Me Improve —Algorithm Aversion and Demonstrating the Ability to Learn. *Business & Information Systems Engineering*, 63(1), 55–68.

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, 181, 21–34.

Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding robots responsible: The elements of machine morality. *Trends in Cognitive Sciences*, 23(5), 365–368.

Biró, P., van de Klundert, J., Manlove, D., Pettersson, W., Andersson, T., Burnapp, L., Chromy, P., Delgado, P., Dworczak, P., Haase, B., Hemke, A., Johnson, R., Klimentova, X., Kuypers, D., Nanni Costa, A., Smeulders, B., Spieksma, F., Valentín, M. O., & Viana, A. (2021). Modelling and optimisation in European kidney exchange programmes. *European Journal of Operational Research*, 291(2), 447–456.

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127–151.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573–1576.

Bostrom, N. (2014). *Superintelligence*. Oxford: Oxford University Press.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin*, 17(5), 475–482.

Brink, K. A., Gray, K., & Wellman, H. M. (2019). Creepiness creeps in: Uncanny valley feelings are acquired in childhood. *Child Development*, 90(4), 1202–1214.

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, 33(2), 220–239.

Cadario, R., Longoni, C., & Morewedge, C. (2021). Understanding, Explaining, and Utilizing Medical Artificial Intelligence. *Nature Human Behaviour*, https://doi.org/10.1038/s41562-021-01146-0.

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

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, 538(7623), 20–23.

Cinelli, M., Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118.

Chouard, T. (2016). The Go files: AI computer clinches victory against Go champion. *Nature*. http://doi.org/10.1038/nature.2016.19553

Common Cents Lab (2017, August 18). Retrieved May 2021, http://advanced-hindsight.com/commoncents-lab/

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, 108(2), 353–380.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571–582.

Decety, J., & Cowell, J. M. (2014). The complex relation between morality and empathy. *Trends in Cognitive Sciences*, 18(7), 337–339.

Diehl, K., Kornish, L. J., & Lynch Jr, J. G. (2003). Smart agents: When lower search costs for quality information increase price sensitivity. *Journal of Consumer Research*, 30(1), 56–71.

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

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

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155–1170.

Dockrill, P. (2017). In just 4 hours, Google's AI mastered all the chess knowledge in history. *Science Alert*.

Donnelly, L. (2017), "Forget your GP, robots will 'soon be able to diagnose more accurately than almost any doctor,' " *The Telegraph*, Retrieved May 19, 2021, from https://www.telegraph.co.uk/technology/2017/03/07/robots-will-soon-able-diagnose-accurately-almost-doctor/

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. *Human Factors*, 44(1), 79-94.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.

Filiz, I., Judek, J. R., Lorenz, M., & Spiwoks, M. (2021). Reducing algorithm aversion through experience. *Journal of Behavioral and Experimental Finance*, 100524.

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349-379.

Fogg, B. J. (2002). *Persuasive technology: Using computers to change what we think and do*. San Francisco: Morgan Kaufmann.

Fournier, T. (2016). Will my next car be a libertarian or a utilitarian?: Who will decide?. *IEEE Technology and Society Magazine*, 35(2), 40-45.

Gogoll, J., & Uhl, M. (2018). Rage against the machine: Automation in the moral domain. *Journal of Behavioral and Experimental Economics*, 74, 97-103.

Goodyear, K., Parasuraman, R., Chernyak, S., Madhavan, P., Deshpande, G., & Krueger, F. (2016). Advice taking from humans and machines: An fMRI and effective connectivity study. *Frontiers in Human Neuroscience*, 10, 542.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619.

Gray, K., Schein, C., & Cameron, C. D. (2017). How to think about emotion and morality: Circles, not arrows. *Current Opinion in Psychology*, 17, 41-46.

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125-130.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), 101-124.

Green, K. C., & Armstrong, J. S. (2007). The ombudsman: Value of expertise for forecasting decisions in conflicts. *Interfaces*, 37(3), 287-299.

Greene, J., Rossi, F., Tasioulas, J., Venable, K., & Williams, B. (2016, March). Embedding ethical principles in collective decision support systems. In *Proceedings of 30th AAAI Conference on Artificial Intelligence* (pp. 4147-4151).

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105-2108.

Grove, W. M., & Lloyd, M. (2006). Meehl' s contribution to clinical versus statistical prediction. *Journal of Abnormal Psychology*, 115(2), 192-194.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814-834.

Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology*, 65(4), 613-628.

Harvey, C. R., Rattray, S., Sinclair, A., & Van Hemert, O. (2017). Man vs. Machine: Comparing Discretionary and Systematic Hedge Fund Performance. *The Journal of Portfolio Management*, 43(4), 55-69.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10(3), 252-264.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, 1(3), 333-342.

Horowitz, M. C. (2016). The ethics & morality of robotic warfare: Assessing the debate over autonomous weapons. *Daedalus*, 145(4), 25-36.

Ireland, L. (2020). Who errs? Algorithm aversion, the source of judicial error, and public support for self-help behaviors. *Journal of Crime and Justice*, 43(2), 174-192.

Ipsos Mori. (2020, February). *Public attitudes towards online targeting* - a report by Ipsos MORI for the Centre for Data Ethics and Innovation and Sciencewise (Research report). Ipsos Mori. Retrieved May 19, 2021, from https://www.ipsos.com/ipsos-mori/en-uk/public-attitudes-towards-online-targeting

Jago, A. S. (2019). Algorithms and authenticity. *Academy of Management Discoveries*, 5(1), 38-56.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.

Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129(3), 339-375.

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its structure, functions, and elective affinities. *Annual Review of Psychology*, 60, 307-337.

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* (pp. 1-16).

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, 93(5), 1449-1475.

Kaya, O., Schildbach, J., AG, D. B., & Schneider, S. (2017, August 10). Robo-advice-a true innovation in asset management. Deutsche Research. Retrieved 2021, https://www.dbresearch.com/PROD/DBR_{INTERNET}*{EN}*-*PROD/PROD0000000000449010/Robo-advice-a*{true}*{innovation}*{in}*{asset}*{managemen}.pdf.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, 446(7138), 908-911.

Komatsu, T. (2016, March). Japanese students apply same moral norms to humans and robot agents: Considering a moral HRI in terms of different cultural and academic backgrounds. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 457-458). IEEE.

Kozyreva, A., Lorenz-Spreen, P., Hertwig, R., Lewandowsky, S., & Herzog, S. M. (2021). Public attitudes towards algorithmic personalization and use of personal data online: Evidence from Germany, Great Britain, and the United States. *Humanities and Social Sciences Communications*, 8(1), 1-11.

Kramer, M. F., Borg, J. S., Conitzer, V., & Sinnott-Armstrong, W. (2018). When do people want AI to make decisions? In *Proceedings of First Annual AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society* (pp. 204-209).

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165, 633-705.

Kuncel, N. R. (2008). Some new (and old) suggestions for improving personnel selection. *Industrial and Organizational Psychology*, 1(3), 343-346.

Leachman, S. A., & Merlino, G. (2017). The final frontier in cancer diagnosis. *Nature*, 542(7639), 36-38.

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243-1270.

Lee, J. D., & Seppelt, B. D. (2006). Human factors and ergonomics in automation design. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (pp. 1570-1596). Hoboken, NJ: Wiley.

Lee, M. K., (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 2053951718756684.

Leyer, M., & Schneider, S. (2019). Me, you or AI? How do we feel about delegation. *Proceedings of the 27th European Conference on Information Systems (ECIS).*

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103.

Lohr, Steve (2016, October 17), "IBM is counting on its bet on Watson, and paying big money for it," *New York Times*, Retrieved May 19, 2021, from https://www.nytimes.com/2016/10/17/technology/ibm-is-counting-on-its-bet-on-watson-and-paying-big-money-for-it.html

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, 46(4), 629–650.

Lourenço, C. J., Dellaert, B. G., & Donkers, B. (2020). Whose algorithm says so: The relationships between type of firm, perceptions of trust and expertise, and the acceptance of financial Robo-advice. *Journal of Interactive Marketing*, 49, 107–124.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, 38(6), 937–947.

Lusardi, A., & Mitchell, O. S. (2011). Financial literacy around the world: an overview. *Journal of Pension Economics & Finance*, 10(4), 497-508.

Malle, B. F. (2016). Integrating robot ethics and machine morality: The study and design of moral competence in robots. *Ethics and Information Technology*, 18(4), 243–256.

Malle, B. F., & Scheutz, M. (2014). Moral competence in social robots. In *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering, Ethics in Science, Technology and Engineering* (pp. 1–6).

Malle, B. F., Scheutz, M., Arnold, T., Voiklis, J., & Cusimano, C. (2015, March). Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 117–124). IEEE.

McFarland, M. (2014). Elon Musk: 'With artificial intelligence we are summoning the demon.' - *The Washington Post*. Retrieved 2021, https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/?utm_{term}=.02d648908751

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis, MN: University of Minnesota Press.

Morewedge, C. K., & Kahneman, D. (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, 14(10), 435-440.

Mori, M. (1970). Bukimi no tani [the uncanny valley]. *Energy*, 7(4), 33–35.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.

Niszczota, P., & Kaszás, D. (2020). Robo-investment aversion. *PLoS ONE*, 15(9), e0239277.

Oliveira, H. G. (2017, September). A survey on intelligent poetry generation: Languages, features, techniques, reutilisation and evaluation. In *Proceedings of the 10th international conference on natural language generation* (pp. 11-20).

O' Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy.* New York: Broadway Books.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, 22(4), 390–409.

Otto, A. S., Clarkson, J. J., & Kardes, F. R. (2016). Decision sidestepping: How the motivation for closure prompts individuals to bypass decision making. *Journal of Personality and Social Psychology*, 111(1), 1-16.

Palmeira, M., & Spassova, G. (2015). Consumer reactions to professionals who use decision aids. *European Journal of Marketing*, 49(3/4), 302–326.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.

Pew Research Center. (2019, January). *Facebook algorithms and personal data* (Research report). Pew Research Center. https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personal-data/

Pezzo, M. V., & Pezzo, S. P. (2006). Physician evaluation after medical errors: does having a computer decision aid help or hurt in hindsight?. *Medical Decision Making*, 26(1), 48-56.

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted?. *Journal of Forecasting*, 36(6), 691–702.

Promberger, M., & Baron, J. (2006). Do patients trust computers?. *Journal of Behavioral Decision Making*, 19(5), 455–468.

Reinecke, M. G., Wilks, M., & Bloom, P. (2021). Developmental changes in perceived moral standing of robots. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

Roberts, S. (2017). Christopher strachey' s nineteen-fifties love machine - *The New Yorker*. Retrieved May 19, 2021, from https://www.newyorker.com/tech/elements/christopher-stracheys-nineteen-fifties-love-machine

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159–169.

Shaffer, V. A., Probst, C. A., Merkle, E. C., Arkes, H. R., & Medow, M. A. (2013). Why do patients derogate physicians who use a computer-based diagnostic support system?. *Medical Decision Making*, 33(1), 108–118.

Shin, K. Y., Lee, J. K., Kang, K. H., Hong, W. G., & Han, C. H. (2019). The current applications and future directions of artificial intelligence for military logistics. *Journal of Digital Contents Society*, 20(12), 2433–2444.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., ··· & Hassabis, D. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815.*

Smith, A. (2018). "Public attitudes toward computer algorithms." Pew Research Center. Retrieved May 19, 2021, from http://www.pewinternet.org/2018/11/16/public-attitudes-toward-computer-algorithms/

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, 27(1), 81–102.

Starke, C., & Lünich, M. (2020). Artificial intelligence for political decision-making in the European Union: Effects on citizens'perceptions of input, throughput, and output legitimacy. *Data & Policy*, 2, e16.

Tetlock, P. E. (2009). *Expert political judgment. How good is it? How can we know?* Princeton: Princeton University Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. (1987). *Rediscovering the social group: A self-categorization theory.* Oxford, UK: Blackwell.

van den Berg, J., Patil, S., & Alterovitz, R. (2017). Motion planning under uncertainty using differential dynamic programming in belief space. In H. I. Christensen & O. Khatib (Eds.), *Robotics Research: The 15th International Symposium ISRR* (pp. 473–490). Cham: Springer International Publishing.

Van Swol, L. M. (2009). The effects of confidence and advisor motives on advice utilization. *Communication Research*, 36(6), 857–873.

Voiklis, J., Kim, B., Cusimano, C., & Malle, B. F. (2016, August). Moral judgments of human vs. robot agents. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 775–780). IEEE.

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219–232.

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

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 32(4), 403–414.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 9(2p2), 1–27.

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources.

International Journal of Human-Computer Studies, 100, 48–54.

Zuckerman, G. (2019). *The man who solved the market: how Jim Simons launched the quant revolution*. London: Penguin Random House.

A three-dimensional motivation model of algorithm aversion ZAHNG Yuyan1, XU Liying2, YU Feng1, DING Xiaojun3, WU Jiahua2, ZHAO Liang4 (1 Department of Psychology, School of Philosophy, Wuhan University, Wuhan 430072, China) (2 Department of Psychology, School of Social Sciences, Tsinghua University, Beijing 100084, China) (3 Department of Philosophy, School of Humanities and Social Science, Xi' an Jiaotong University) (4 Department of Publishing Science, School of Information Management, Wuhan University, Wuhan 430072, China) Abstract: Algorithm aversion refers to the phenomenon of people preferring human decisions but being reluctant to use superior algorithm decisions. The three-dimensional motivational model of algorithm aversion identifies three primary causes: doubt regarding algorithmic agency, lack of moral standing, and the annihilation of human uniqueness. These correspond respectively to three psychological motivations—trust, responsibility, and control—and suggest three feasible approaches to mitigate algorithm aversion: enhancing human trust in algorithms, strengthening algorithmic agency responsibility, and exploring personalized algorithm design to highlight human control over algorithmic decisions. Future research should further investigate the boundary conditions of algorithm aversion and other potential motivations from a more social perspective.

Key words: algorithmic decision-making, algorithm aversion, psychological motivation, human-robot interaction

Note: Figure translations are in progress. See original paper for figures.