UNDERSTANDING AFFECTIVE TRUST IN AI:

THE EFFECTS OF

PERCEIVED BENEVOLENCE

By

MARJORIE A. ERDMANN, M.S.

Bachelor of Science in Sociology
DePauw University
Greencastle, Indiana
1991

Master of Science in Family Relations and
Child Development, Marriage and Family Therapy
Oklahoma State University
Stillwater, Oklahoma
December, 1994

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2022

UNDERSTANDING AFFECTIVE TRUST IN AI:

THE EFFECTS OF

PERCEIVED BENEVOLENCE

Dissertation Approved:

Bryan D. Edwards, PhD

Dissertation Adviser

James M. Pappas, PhD

Dursun Delen, PhD

Denna L. Wheeler, PhD

ACKNOWLEDGMENTS

I dedicate my dissertation to my husband, Ken, and my children, Mary and Henry. I thank you for your endless patience, support, and understanding of my time away from you while I toiled away in "the tower" in pursuit of this Ph.D. A special thanks to Mary who *graciously* attended the same university at the same time as her Mom. You always made me feel welcome. A special thanks to Henry and Ken who had to pick up my slack at home. I am so fortunate to have each of you.

I thank my parents and lament them not being here with me to celebrate. They valued education, critical thinking and lively debates. My dad often said that the most important thing to learn is how to think. That was the impetus for me pursuing my Ph.D., to learn to think like the academics I worked with at Oklahoma State University (OSU). A special thanks to my Mom, who encouraged me and celebrated with me at the end of each semester. How I wish I could celebrate finishing this final semester with her.

I thank Andrew Urich, OSU Eastin Center Chair for Talent Development, and longtime friend for advocating for me to be hired by OSU in 2014, advocating for my acceptance into the Executive Ph.D. program in 2019 and giving me an opportunity to teach this year. Your support has changed my career trajectory for which I am deeply grateful.

I thank my superstar dissertation committee, Drs. Bryan Edwards, James Pappas, Dursun Delen and Denna Wheeler. I had the unique advantage of receiving your mentorship through this dissertation process and our work collaborations. I continue to be inspired by and learn from you all. I look forward to many more collaborations.

To Dr. William Paiva and Dean Ken Eastman, I thank you for your investment in my education. Without it, I absolutely would not have been able to participate in the program. You made it possible.

To the best EPhD cohort ever, Cohort 8, I appreciate our comradery, how fun you all made residencies, and how supportive we were of each other. I give a deep, heartfelt thanks to my fellow Rosies—Abbey and Reba; *it is no understatement to say that I could not have succeeded in this program without Rosie power*.

Finally, I want to thank my co-workers, friends, and extended family for listening to my school stories for years now. You stuck with me—I cannot wait to celebrate together.

Name: MARJORIE A. ERDMANN

Date of Degree: DECEMBER, 2022

Title of Study: UNDERSTANDING AFFECTIVE TRUST IN AI: THE EFFECTS OF PERCEIVED BENEVOLENCE

Major Field: BUSINESS ADMINISTRATION

Abstract: The primary objective of this research was to gain understanding of affective trust in AI (how comfortable individuals feel with various AI applications). This dissertation tested a model for affective trust in AI grounded in interpersonal trust theories with a focus on the effects of perceived benevolence of AI—an overlooked factor in AI trust research. In Study 1a, online survey participants evaluated 20 AI applications with single-item measures. In Study 1b, four AI applications were evaluated with multi-item measures. Perceived benevolence was significantly, positively associated with affective trust over and above cognitive trust and familiarity in 21 of 24 AI tests. Confirmatory factor analysis suggested four factors, supporting the theory that cognitive trust and affective trust in AI are distinct factors. The secondary objective was to test the utility of manipulating perceived benevolence of AI. In Study 2, online survey participants were randomly assigned to one of two groups with 10 AI applications described as "augmented intelligence" that "collaborates with" a specific or exact same AI described as "artificial intelligence." The augmentation manipulation did not matter; there were no significant direct or indirect effects to benevolence or affective trust. These results imply that "Augmented Intelligence" positioning has no significant effect on affective trust, counter to practitioners' beliefs. In Study 3, online survey participants were randomly assigned to one of two groups—one that received benevolence messaging (a message informing the participant that the AI was intended for human welfare) for five AI applications and the other did not.

Benevolence messaging was also tested to see if it moderated contexts expected to diminish affective trust (likelihood of worker replacement and likelihood of death from error). Benevolence was not influenced by the manipulation. Surprisingly, likelihood of worker replacement had no significant association with affective trust, and likelihood of death from error had only one significant association. People may be more ambivalent about these contexts than previously thought. Understanding affective trust in AI was expanded by identifying the importance of perceived benevolence. Until benevolence messaging can boost perceptions of benevolence, the success of that strategy remains unknown.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF GRAPHS

CHAPTER I

INTRODUCTION

Artificial intelligence (AI) is expected to transform work and solve societal ills with its awesome computational power. But to do so, humans must be willing to trust it—and research has found people to be reluctant (Berger, Adam, Rühr & Benlian, 2021; Burton, Stein & Jensen, 2020; Castelo, Bos & Lehmann, 2019; Dietvorst, Simmons & Massey, 2015, 2018; Dzieżyk & Hetmańczuk, 2020; Logg, Minson & Moore, 2019; Prahl & Van Swol, 2017; Yeomans, Shah, Mullainathan & Kleinberg, 2019). This reluctance to trust AI threatens to delay our ability to harness AI's power to benefit individuals, firms, and society (Huang & Rust, 2018). Reluctance to trust AI has already delayed the adoption of effective AI in medicine (Reis, Maier, Mattke et al., 2020), in commercial real estate (McGrath, Desai & Junquera, 2019), in employee recruitment and selection (Ore & Sposato, 2021), and in criminal justice (Bagaric, Hunter & Stobbs, 2019). The reluctance is partly fueled by people's negative attitudes toward AI (Schepman & Rodway, 2020; Zhang & Dafoe, 2019). People are uncomfortable with AI for several reasons, including concern with its use for traditionally intuitive decisions (Castelo et al., 2019; Yeomans et al., 2019), disappointment at learning that AI is imperfect (Dietvorst et al., 2015; Glikson & Woolley, 2020; Madhavan & Wiegmann, 2007; Prahl & Van Swol, 2017), worry about AI replacing human workers (Huang & Rust, 2018; Schepman & Rodway, 2020), and apprehension about its application in life and death scenarios (Schepman & Rodway, 2020). The general negativity toward AI is so pervasive that medical and automotive companies often avoid describing their AI products with the terms "AI" or "artificial intelligence" (Hengstler, Enkel & Duelli, 2016). Therefore, to harness the promise of AI, it is critical to understand the bases of comfort with artificial intelligence.

Thus far, scholars have focused on increasing trust in AI by explaining how accurate or *good at* a task AI is, relying on performance to generate trust. The results of these efforts are mixed. For example, Castelo et al. (2019) explained to research participants that AI algorithms outperformed qualified humans and, in another study, explained that AI algorithms can effectively solve intuitive tasks using data. Both performance explanations increased perceptions of AI effectiveness, but both resulted in participant ambivalence toward using the algorithms, falling short of evoking trust in AI. Dzieżyk and Hetmańczuk (2020) provided participants with feedback on the algorithm's excellent performance, but feedback did not improve participants' decisions to use the AI. Dietvorst et al. (2015) found resistance to using AI even when the person witnessed first-hand that the AI outperformed a qualified human, a phenomenon coined "algorithm aversion." In short, relying on AI's performance has not elicited a threshold of trust to garner use. Because a good reason to trust something (i.e., superior performance) is considered a basis for cognitive trust (vs. affective trust) (McAllister, 1995), it is fair to summarize many tactics to date as targeting cognitive trust.

A primary assumption has been that people who learn about AI's superior performance will also be more comfortable with using AI. Interestingly, a study conducted by Castelo et al (2019) revealed that performance and comfort are not necessarily linked. Castelo et al. (2019) used a *cognitive trust tactic* (explained to participants that certain human judgement tasks are best solved using quantifiable personality traits) which increased cognitive trust; however, the manipulation *had no effect on participants' comfort* with the AI applications. They operationalized affective trust via self-reported discomfort. Castelo et al.'s results revealed that AI cognitive trust tactics may not influence AI affective trust (i.e., discomfort), which is a significant departure from the affective trust theory whereby affective trust is directly and strongly influenced by cognitive trust (Johnson & Grayson, 2005; Lewis & Weigert, 1985; McAllister, 1995; Schaubroeck, Lam & Peng, 2011). It is possible that affective trust in AI is an exception. If so, this may help explain why resistance to AI persists even when performance is exceptional. Given that interventions targeted towards cognitive trust have no effect on affective trust in AI, I argue that learning how to directly influence affective trust in AI could be critical to reaching a trust threshold high enough to elicit the eventual acceptance and adoption of AI.

Three conditions have shown to improve comfort in AI: (a) when the AI has human characteristics (de Visser, Monfort, McKendrick et al., 2016); (b) when people are familiar with AI (Belanche, Casaló & Flavián, 2019; Castelo et al., 2019); and (c) when people have control over AI (Dietvorst et al., 2018; Zlotowski, Yogeeswaran & Bartneck, 2017). Nevertheless, fulfilling each of these conditions comes with other unintended problems. Human-likeness can also be eerie to people which causes discomfort (Mende, Scott, van Doorn et al., 2019). Familiarity works when the experience with AI is positive, which may not always be the case (Berger et al., 2021). Control works but relinquishing control to humans could mute AI benefits and undermine the reason to deploy AI in the first place (Burton et al., 2020; Möhlmann & Henfridsson, 2019; Mohlmann, Zalmanson, Henfridsson & Gregory, 2021; Muir, 1987).

If superior performance does not increase affective trust, if humanization and familiarity can both increase or decrease affective trust, and if developers and implementers want to avoid relinquishing control of AI to users, the question remains, "What could reliably increase affective trust in AI?"

I propose that the answer lies in the interpersonal trust theory (Mayer, Davis & Schoorman, 1995; McAllister, 1995), which highlights a gap in current human-AI trust research (Glikson & Woolley, 2020). In human-to-human trust, perceiving another as being interested in and acting in your general welfare, a construct called benevolence is a basis of trust (Mayer et al., 1995; Schoorman, Mayer & Davis, 2007; Wang & Benbasat, 2007) that has been found to be associated with affective trust (McAllister, 1995). Even more encouraging for the potential importance of benevolence perceptions are these three factors: (1) benevolence is something that is signaled to another party and so benevolence signaling (via messages) could provide marketers and managers with a strategy to promote affective trust; (2) benevolence seems unlike other AI affective trust antecedents in that it may reliably increase trust given that I found no studies of benevolence perceptions lowering trust (in humans); and (3) it may help overcome early reluctance to use AI because initial trust is a predictor of technology use (Li, Hess & Valacich, 2008). Affective trust is particularly important at the beginning of relationships at the time of initial trust formation (McAllister, 1995; Schoorman et al., 2007), when AI aversion most often occurs (Dietvorst et al, 2015).

Contribution to Literature

Four studies were conducted to: (a) test the notion that people evaluate benevolence of AI whereas I measured the association between benevolence perceptions and affective trust while controlling for familiarity and cognitive trust; (b) test if perceived benevolence helps explain why algorithms described as augmenting human workers, defined as working collaboratively with a worker (Raisch & Krawkowski, 2021) was associated with higher levels of affective trust; and (c) test if benevolence was malleable by using benevolence messaging to test if it moderated the relationship between two common concerns with AI (worker replacement and death from error) and affective trust. My studies generally follow Castelo et al.'s (2019) progression in their research that created a cascading argument about perceived objectivity of the task mattering to AI trust. I leveraged previous AI attitude research results to identify algorithms high and low in comfort to use in the studies.

The current study contributes to human-AI trust theory by exploring perceived benevolence's association with affective trust and developing a model for affective trust in AI. This study supports the use of the human-human trust theory to identify a new humanizing strategy (signaling benevolence) for AI trust, thus expanding on AI humanization research. This study examines assumptions about affective trust in augmenting AI as well as a practical strategy—benevolence messaging—that marketers (who are tasked with selling AI tools) and managers (who are tasked with implementing AI tools for the company's benefit) could use to increase consumers' and employees' affective trust in artificial intelligence.

CHAPTER II

REVIEW OF LITERATURE AND HYPOTHESIS DEVELOPMENT

Artificial intelligence (AI) is defined in the management field as "…a new generation of technologies capable of interacting with the environment by (a) gathering information from outside (including from natural language) or from other computer systems; (b) interpreting this information, recognizing patterns, inducing rules, or predicting events; (c) generating results, answering questions, or giving instructions to other systems; and (d) evaluating the results of their actions and improving their decision systems to achieve specific objectives" (Glikson & Woolley, 2020, p. 628). Although some of the applications are robotic, others are algorithmic. Algorithms are "a set of steps that a computer can follow to perform a task" (Castelo et al., 2019, p. 809) and are the result of analyzing and learning from existing data (Dietvorst et al., 2018). AI has proven capable of outperforming human beings, including experts, at many tasks, creating optimism about its potential benefits to humans (Castelo et al., 2019; Fan, Liu, Zhu & Pardalos, 2020; Zhang & Dafoe, 2019).

However, achieving those benefits is challenged by human resistance to AI adoption. An extensive literature review of AI adoption research by Glikson and Woolley (2020) found that a significant barrier to AI usage was a lack of trust in AI. Therefore, understanding how to build trust in AI is generating significant interest among scholars and developers because trust in technologies and automations have reliably predicted their use (Glikson & Woolley, 2020; Hoff & Bashir, 2015; Lee & See, 2004; Madhavan & Wiegmann, 2007). As a result, fostering trust in AI is considered critical to organizations' eventual ability to capture gains from AI implementations and companies' ability to sell AI products (Glikson & Woolley, 2020; Huang & Rust, 2018).

In this current study, I embrace the existence of two distinct types of trust: cognitive trust (rational thoughts) and affective trust (feelings) (McAllister, 1995). I argue that there are three significant gaps in AI trust research. First, AI researchers have focused on intentionally building cognitive trust by highlighting AI's superior abilities to analyze data while paying comparatively minimal attention to intentionally building affective trust in AI. Second, due to a lack of focus on affective trust, we know very little about it (Glikson & Woolley, 2020). What little research we have comes from robot-human interaction studies measuring how robot features and actions make people feel. The minimal insight we have into affective trust in AI is especially problematic for organizations because we know the least about the type of AI they desire to implement—algorithmic AI for decision aids, virtual AI for chat box assistants, or embedded algorithms for cuing other computer systems. Third, there has been little validation of strategies that boost affective trust at the time of initial trust formation even though research has shown that negative feelings toward AI could be leading people to not trust or use it—even when its performance is exceptional and exceeds human performance (Dietvorst et al., 2015).

In this dissertation, I integrate two human-human trust theories and propose a model for affective trust. Then, I apply that model to affective trust in AI. Other human-human trust theories have been successfully applied to explain human trust (but not affective trust) in technologies, computers, and automation (Hoff & Bashir, 2015; Lee & See, 2004; Madhavan & Wiegmann, 2007; Muir, 1987), giving me confidence in the utility of the human-human trust theory to identify gaps and highlight opportunities for affective trust in AI. Two leading human trust theories (Mayer et al., 1995; McAllister, 1995) were reviewed, highlighting that benevolence, "the extent to which a party is believed to want to do good for the trusting party, aside from an egocentric profit motive" (Mayer et al, 1995, p. 345), is an overlapping concept and a trust base, specifically an affective trust base. The review of literature includes current AI attitude and adoption research identifying evidence of benevolence perceptions as an affective trust base in AI.

Integration of Benevolence as an Affective Trust Antecedent

In 1995, trust had become a central topic for organizational researchers seeking to explain interpersonal relationships at work and risk-taking behavior (Mayer et al., 1995) and explore worker and manager performance (McAllister, 1995). Two theories of interpersonal trust emerged as the most influential. First, McAllister (1995) published an empirical article proposing and supporting a trust theory with affective and cognitive "foundations" for trust (p. 24). McAllister defined interpersonal trust as "the extent to which a person is confident in, and willing to act on the basis of, the words, actions, and decisions of another" (p. 25.). McAllister embraced Johnson-George and Swap's (1982) two dimensions of trust—"reliableness" and "emotional trust." McAllister explained that "Trust is cognition-based in that 'we choose whom we will trust in which respects and under what circumstances, and we base the choice on what we take to be 'good reasons,' constituting evidence of trustworthiness" (Lewis & Weigert, 1985 cited in McAllister (1995, p. 25) whereas, affect-based trust comes from trying to determine how another's actions "will affect him or her" (p. 25). Ultimately, when deciding to trust another, people ask themselves, "Do they consider my interest and welfare?" (p. 25). The answer to this question creates a reassurance, or a comfort level referred to as affective trust.

In the same year, Mayer, Davis and Schoorman published a competing theoretical article proposing a trust theory with ability, integrity, and benevolence as "factors of perceived trustworthiness" (Mayer et al., 1995, p. 719). Mayer and colleages defined trust as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the actor, irrespective of the ability to monitor or control that other party" (p. 712). Mayer et al. did not emphasize but did recognize that an "affective link with a trustee" (p. 725) can result from a relationship where people are each taking risks.

Although Schoorman et al. (2007) acknowledged research showing that affect has an influence on trust [e.g., emotional state research by Jones and George (1998) and Dunn and Schweitzer (2005)], they ultimately discounted affective trust and explained that it was still their contention that trust is solely a result of rational thinking (cognitive trust). Nevertheless, I argue that Mayer's trust theory is very similar to McAllister's when considering Mayer et al.'s benevolence trust base. I argue benevolence is conceptually congruent with McAllister's description of affective trust bases, creating an integration point for the two theories. I argue that perceptions of Mayer et al.'s benevolence, according to its definition and description of how it is detected, is McAllister's basis of affective trust. First, consider the following description of benevolence.

In Schoorman et al.'s (2007) study, benevolence was defined as "the extent to which a party is believed to want to do good for the trusting party, aside from an egocentric profit motive" (p. 345), and is an evaluation of the other party's intentions and motives and the conclusion that actions are based on wanting to be helpful when it is not required to be helpful and when "there is not extrinsic reward for the mentor" (Mayer et al., 1995, p. 719). According to Mayer et al. (1995), benevolence "is the perception of a positive orientation of the trustee toward the trustor" (p. 719), while benevolence actions are associated with "strong bonds" (p. 345) between parties.

Consider the overlap of this description of benevolence with McAllister's (1995) description of bases of affective trust (see Table 1). Affective foundations for trust also exist, consisting of emotional bonds between individuals (Lewis & Wiegert, 1985). People make emotional investments in relationships, express genuine concern and care for the welfare of partners…" (McAllister, 1995, p. 26). "Insights into the motives of relationship partners provide foundations for affect-based trust" (McAllister, 1995, p. 29). Because organizational citizenship behavior is behavior “intended to provide help and assistance that is outside an individual’s work role, not directly rewarded…” (McAllister, 1995, p. 29), the level of organizational citizenship behavior is associated with affective trust levels.

Table 1

Benevolence Commonality Table

Common Themes	**Mayer et al.’s (1995) Benevolence**	**McAllister’s (1995) Affective Bases**
Trustor’s general welfare	“want to do good for the trusting party, aside from an egocentric profit motive” (p. 345)	“express genuine concern and care for the welfare of partners…” (p. 26)
Helpful when not required	“wanting to be helpful when it is not required to be helpful and when “there is not extrinsic reward for the mentor” (p. 719)	“intended to provide help and assistance that is outside an individual’s work role, not directly rewarded…” (p. 29)
Evaluation based on perceived intentions and motives	“evaluation of the other party’s intentions and motives” (p. 719)	“Insights into the motives of relationship partners provide foundations for affect-based trust” (p. 29)
Bonds	“strong bonds” (p. 345)	“emotional bonds between individuals” (p. 26)

Clearly, what Mayer et al. (1995) and McAllister (1995) describe is a very similar base to overall trust, whereas Mayer et al. went the extra step and named benevolence. Therefore, for this current study, I argue that benevolence is a basis of affective trust.

Further supporting benevolence as an affective base, my review of organizational literature for affective trust bases identified five constructs which were identified as empirically, positively associated with affective trust. They are cognitive trust, familiarity, citizenship behavior, servant leadership, and good reputation.

(1) *Cognitive trust* (Johnson & Grayson, 2005; Lewis & Weigert, 1985; McAllister, 1995; Schaubroeck et al., 2011). Lewis and Wiegert (1985) explained that with cognitive trust, “we choose whom we will trust… ... [and] base the choice on what we take to be ‘good reasons,’ constituting evidence of trustworthiness” (p. 970). In all four trust models

reviewed, findings show that cognitive trust had the strongest association with affective trust.

(2) *Familiarity* (Johnson & Grayson, 2005; Lewis & Weigert, 1985; McAllister, 1995). Frequent, satisfying interaction provides a sufficiency of data to have confidence in another's attributions.

(3) *Citizenship behavior* (McAllister, 1995). Behavior demonstrating an interest in the care and concern of another outside of work roles was associated with higher affective trust.

(4) *Servant leadership* (Schaubroeck et al., 2011). Leadership behavior "emphasizes promoting the welfare of others" (p. 865) which results in higher affective trust.

(5) *Good reputation* (Johnson & Grayson, 2005). Affective trust results when firms do things that are right and are "an expression of empathy for the customer" (p. 502).

I propose that the last three antecedents listed (citizenship behavior, servant leadership, and good reputation) share a common explanation that perceiving the other party as interested in what is *good for the trustor*, as oriented toward the trustor, and as acting in the trustor's general welfare. I argue that these have overlap to Mayer et al.'s (1995) benevolence and based on the mechanisms of action used to explain the three constructs' association with affective trust, benevolence is the common factor. Obviously, AI cannot signal benevolence via citizenship behavior or leadership because these are human qualities. However, benevolence perceptions can be communicated in other ways as a characteristic of the AI application. Thus, within the context of AI, there would be three bases of affective trust: perceived benevolence, cognitive trust, and familiarity (see Figure 1). This integration supports Mayer et al.'s defense of benevolence as a relevant and necessary trust base. Mayer and colleagues point out that perceived benevolence of another has been part of interpersonal trust theories for decades (i.e., Larzerlere & Huston, 1980; Solomon, 1960; Strickland, 1958 cited in Mayer et al., 1995).

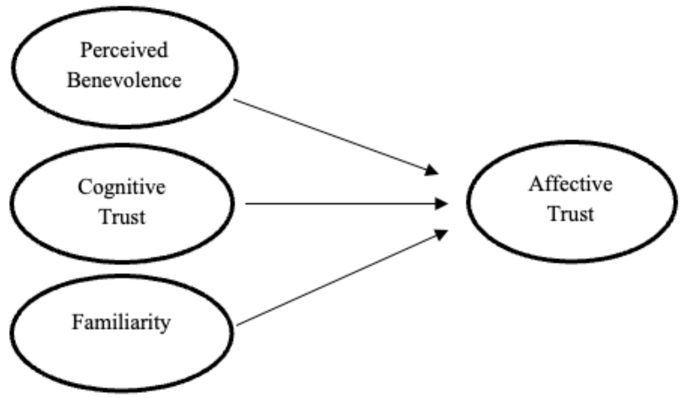


Figure 1. Conceptual model for affective trust bases.

Affective Trust in AI

Artificial intelligence acceptance and adoption researchers from marketing and computer science agree that there is both cognitive and affective trust in AI. For example, attitude researchers have measured human perceptions of AI's capabilities (Schepman & Rodway, 2020) or effectiveness of AI (Castelo et al., 2019) as their measures for cognitive trust. They have also examined (dis)comfort for their measure of affective trust (Castelo et al., 2019; Schepman & Rodway, 2020). They argue that comfort is a feeling one has about the AI and that emotion is a proxy for affective trust levels. For the purposes of the present research and in alignment with previous AI attitude research, I embrace that participant ratings of comfort in AI as one way to measure affective trust.

Glikson and Woolley's (2020) theory of human trust in AI was derived from McAllister's (1995) affect- and cognition- based trust theory. Nevertheless, their affective trust bases for human-AI trust departed from organizational literature and instead were formed in response to research findings. They grouped common research together and named constructs associated with human-AI cognitive trust and human-AI emotional trust. Naming types of trust bases helped them to achieve their goal of integrating findings from across multiple disciplines such as computer science, automation and human factors. Three human-AI affective trust bases (not linked to interpersonal affective trust theories) were identified: tangibility (physical presence, persona), anthropomorphism (human-likeness, attractiveness, facial similarity), and immediacy behaviors (expressiveness, conversationality, praise) (Glikson & Woolley, 2020).

Glikson and Woolley's (2020) extensive review revealed two major gaps in affective trust in AI research. First, because AI trust research has not directly targeted affective trust—especially with non-robotic AI—we know very little about it. What is known about affective research is regarding human responses to robots rather than AI that is virtual like chat boxes or embedded within systems like algorithms. Because Glikson and Woolley's (2020) approach to building their affective trust framework was reactive to current research, robotic features had a heavy influence on the dimensions identified. Thus, all of the dimensions rely heavily on a physical presence, visible presence onscreen or persona. For example, people are more comfortable approaching sitting robots versus standing robots (Obaid, Sandoval, Zlotowski et al., 2016). People like robots who give interpersonal signals like nodding to indicate listening (Jung, Lee, DePalma et al., 2013). People respond more positively to chat-boxes with an avatar and a persona (Chatterman, Kwon, Gilbert & Li, 2014). Depending on physical presence is not informative for algorithmic AI and embedded AI that is used in most organizational AI tools because embedded AI is not seen like a robot but works behind the scenes via computer processing. However, because the robot trust signals are so congruent with signals for human-human trust (sitting is less threatening, nodding signals engagement) gives confidence to the idea that human-human trust theories meaningfully inform how to build trust in AI conditions. Afterall, we have humanized AI—maybe unintentionally; we refer to it as intelligent, and promote its ability to learn. This leads to the possibility that other human characteristics like benevolence will signal trustworthiness. The advantage that benevolence has is without a dependence on physical presence.

Second, cues for affective trust identified by Glikson and Woolley (2020) seem vulnerable to human interpretation or contexts, meaning designers may not be able to rely upon them to induce affective trust. For example, anthropomorphism may or may not evoke affective trust. According to Glikson and Woolley (2020), “Human-likeness mostly increases positive emotions, but can also cause discomfort” (p. 643). Tangibility may or may not evoke affective trust. “Physical presence may not only increase liking but also induce fear” (Glikson & Woolley, 2020, p. 643). Benevolence perceptions, on the other hand, maybe more stable. I found no research where benevolence was not positively associated with trust.

The gaps in the affective trust bases identified by Glikson and Woolley suggest that there are yet unidentified influential evaluations of AI that should be explored. The current AI attitudinal research provides clues that the evaluations taking place consist of my proposed human affective trust bases (familiarity, performance/cognitive trust, and benevolence), with particularly high relevance of benevolence perceptions for AI.

Benevolence

Schepman and Rodway (2020) examined AI attitudes by surveying people in two different ways. First, they surveyed people’s agreement with positive and negative statements about AI. The three top (of 16) rated positive statements were: (1) “There are many beneficial applications of Artificial Intelligence;” (2) “I am impressed by what Artificial Intelligence can do;” and (3) “Artificial Intelligence can have positive impacts on people’s wellbeing.” Therefore, it is reasonable to conclude that positive feelings about specific AI are derived from knowledge of many beneficial applications (familiarity), the impressiveness of performance (cognitive trust), and being in people’s well-being (benevolence), supporting the integrated affective trust model with three bases for affective trust as proposed in this dissertation.

Interestingly, the three top (of 16) rated negative statements were: (1) “The rise of Artificial Intelligence poses a threat to people’s job security,” (2) “I am concerned about Artificial Intelligence applications mining my personal data,” and (3) “Artificially intelligent systems should be banned from making life or death decisions." All three statements reflect conditions where AI use would not be for the general welfare of humans: job loss, privacy violation, and death. I argue these reasons point to the weight of non-benevolent perceptions of AI. None of the three mention a lack of familiarity or concern about performance superiority (cognitive trust) as a reason to be aversive to AI. Based on this evidence, I propose that low benevolence perceptions are the leading reason for negative attitudes toward AI, which makes benevolence perceptions especially relevant.

Schepman and Rodway (2020) also asked people to rate the capability of and their comfort in 42 AI applications that they collected from public news stories. They concluded that both capability and comfort was reflective of whether the tasks "involve(ed) big data/automation" or if the tasks "involve(ed) human judgement." Involvement of Big Data may be influential, but I argue that the different comfort levels may have also been influenced by the participants' benevolence evaluations, the extent to which the description expressed a benefit for human welfare. Consider Schepman and Rodway's (2020) AI descriptions with regards to whether they communicate algorithms' orientation toward human general welfare or not. Two AI application descriptions associated with high comfort were "Using smells in human breath to detect illness" and "Translating speech into different languages in real time." Detecting illness and facilitating communication are clearly in humans' best interests. Compare those with two AI application descriptions that were associated with low comfort: "Driving a car" and "Being a bank branch employee." I argue that driving a car and being a bank employee are descriptions which lack any cues about how they are intended for human benefit. I can imagine research participants feeling comforted by applications described as providing a beneficial role or being in humans' best interest and feeling less comfortable with those described with little to no cues related to general human welfare.

In addition to the potential cues about benevolence in Schepman and Rodway's survey research, there is further indirect evidence that people make benevolent evaluations of AI from AI use research: use seems associated with conditions where people conclude that the AI is in their general welfare. For example, Longoni, Bonezzi and Morewedge (2019) tested people's preference for medical advice (regarding skin cancer screenings, implantation of pacemaker, and emergency triage) from a human versus a computer (AI) and concluded that people preferred human advice due to "uniqueness neglect" from AI. However, Pezzo and Beckstead (2020) re-analyzed Longoni et al.'s (2019) data and concluded that people preferred AI in all the conditions when the medical advice from AI was more accurate than human doctors, in contrast to Longoni et al.'s (2019) interpretation of general aversion to AI. Pezzo and Beckstead's analysis showing adoption of AI once it was superior to humans departs from a bulk of other AI usage research where superiority was not enough to entice trust and use. I argue that the difference in Longoni et al.'s (2019) study is that the superiority condition of the medical advice AI is confounded with also changing the benevolence evaluation of the AI—as better care is in the person's best interest (health).

Conceiving that the difference was in participants' benevolence evaluation helps explain why the adoption pattern was different than other research and why, in this case, superiority translated to use. I argue that the benevolence evaluation of better care provided the affective trust boost needed to overcome their aversion to using the AI; notably, aversion that was present when the AI performed the same as a human doctor. When the AI performed the same, there was a cognitive reason to trust it (it performed as well as a qualified doctor) but there was not a benevolence advantage to doing so, leaving patient trust below the threshold to induce use. How people felt about the AI providing medical advice was not captured so it is unknown what degree of trust in the AI came from cognitive or affective sources. Hence, studies aimed at untangling cognitive trust from affective trust are needed to illuminate the effects and barriers of each to better explain use decisions to, in turn, address reluctance.

Cognitive Trust

Researchers have attempted to increase trust in AI using cognitive trust or 'good reasons' to trust AI, but with minimal success. For example, some researchers have claimed that the culprit is the nature of the task itself, explaining that people trust AI when the task is data-based and do not trust AI when the task is human judgement or intuitive-based (Castelo et al., 2019; Schepman & Rodway, 2020). Castelo and colleagues (2019) attempted to boost trust by giving participants evidence of AI's superior performance for subjective tasks, which boosted trust but not enough to confidently conclude that performance data alone would overcome AI aversion.

The most compelling evidence for a weak relationship existing between cognitive trust and affective trust and thus increasing interest in benevolence perceptions was from one of Castelo et al.'s (2019) manipulations. They found an anomaly about affective trust in AI. Explaining to participants that quantifiable data could be used to make human judgements for recommending a movie and recommending a romantic partner and therefore positioning AI as effective for the tasks did, in fact, increase participant perceptions about AI effectiveness. Essentially, telling people AI was effective increased the degree to which the participants rated it as effective. Comfort ratings (Castelo et al.'s measure for affective trust) for the AI applications, however, was unaffected.

Contrary to affective trust theory, cognitive trust did not boost affective trust. If affective trust in AI is not as influenced by cognitive trust, this might help explain why aversion has been relatively immune to cognitive reasons. This anomaly raises interest in the other affective trust bases. We know that, theoretically, cognitive trust is important to affective trust (Johnson & Grayson, 2005; Lewis & Weigert, 1985; McAllister, 1995; Schaubroeck et al., 2011), but we do not know the strength of the relationship for AI.

Familiarity

Scholars have found that people's familiarity with the AI application increases trust (Castelo et al., 2019) and use (Belanche et al., 2019; Logg et al., 2019). In Logg and colleagues'

(2019) research, they found that their participants were mostly familiar with AI by measuring if they could choose the accurate definition of AI. They found that participants also appreciated algorithmic advice. According to Berger et al. (2021), as people became more familiar with an AI application where the AI is learned and improved, participants became less aversive to AI. Algorithmic literacy programs to increase familiarity with AI have been proposed as a countermeasure to aversion (Burton et al., 2020; Musen, Middleton & Greenes, 2014). Literacy programs for AI support what we also know from organizational research where familiarity has been associated with affective trust (Johnson & Grayson, 2005; Lewis & Weigert, 1985; McAllister, 1995).

In summary, given my theoretical argument for benevolence as a basis of affective trust and given the support from AI research regarding the possible effects of benevolence, I hypothesize that perceived benevolence of AI will be associated with affective trust. Because cognitive trust theoretically is associated with affective trust (even though it is possibly not strongly linked in AI) and because familiarity is a known factor associated with affective trust, I control for those in determining the strength of association between benevolence perceptions and affective trust in AI (see Figure 2).

H1: *Perceived benevolence will be positively associated with affective trust in AI while controlling for cognitive trust and familiarity.*

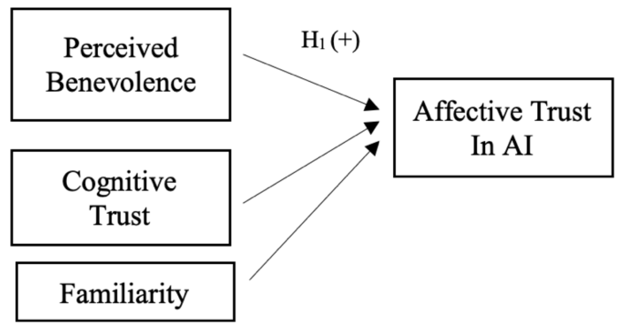


Figure 2. Perceived benevolence hypothesized relationship to affective trust with controls

Benevolence of Augmented AI

Augmentation Role

Understanding the importance of benevolence perceptions could help explain people's proclivity for AI intended to augment a worker even though AI under the control of a human may be less effective (Muir, 1987). Researchers have encouraged developers to focus on AI for augmenting a worker as a strategy to gain acceptance (Burton et al., 2020). Physicians are an excellent example of a profession embracing that strategy. The *Journal of the American Medical Association* (JAMA) specifies that "AI" in JAMA standards stands for "augmented intelligence," meant to assist human decisions, not replace human decisions (American Medical Association, 2022). But what features of the augmented context create comfort?

I propose that AI described with an augmenting role likely triggers a sense that it is benevolent and thus, AI for augmentation is associated with higher affective trust. Burton et al. (2020) speculated that augmentation comforts people by making them feel "in the loop." It is logical that when an algorithm's role is described as collaborative, or when humans are "in the loop," that is a way to signal that the AI is benevolent, intended for their best interest (e.g., to improve work performance). And we know from organizational research that benevolence is something that is signaled to the trustor (e.g., servant leadership).

Schepman and Rodway's (2020) survey did not explicitly define whether AI was augmenting or not, but it may have been signaled by some descriptions such as "helping farmers remove weeds and collect the harvest" (p. 8). When AI was described as contributing to a worker by reducing a workload (e.g., "helping investment bankers make decisions modeling different scenarios") or by detecting something difficult for a worker to detect (e.g., "helping detect life on other planets") (p. 8), these AI types fell among those associated with higher comfort. However, we do not know if intentionally describing AI as augmenting a worker affects initial attitudes about the AI.

After reviewing AI usage research, I concluded that the positive emotional effect of augmentation has potentially been established, albeit unintentionally, through usage manipulations that may have signaled to the user the benevolent role of the AI. For example, Dietvorst et al. (2018) signaled that the AI was benevolent by giving participants some control over it. Logg et al. (2019) signaled that AI was benevolent by allowing participants to collaborate with the AI, allowing them to adjust predictions after AI feedback. In each case, when AI was augmented, the participants' use of AI was higher.

Control as a Benevolence Signal

Evidence that control's effect is through benevolence perceptions was found in Dietvorst et al.'s (2018) research where the amount of control varied (in one condition, the participants could make no adjustments; in another condition, the participants could adjust the prediction by up to 10 percentiles; and in another condition, participants could adjust the predictions as much as they liked). Surprisingly, the amount of control did not change the degree to which people felt satisfied with or used the AI. Dietvorst et al. (2018) determined that controlling AI made it "much more palatable" (p. 1168). I argue that offering people control of the AI may have signaled its benevolence and I propose that the effect of control on use (hence trust) was met through benevolence perceptions rather than through a sense of agency since control levels themselves did not translate into differing levels of satisfaction or use.

Allowing people to control AI may trigger benevolence perceptions, but it is also fraught with problems; it may not generate the outcomes intended by AI usage. Research has found that in some cases, when people gain control of AI, they may not use the AI for its organizational intent. For example, Uber drivers who understood how to control incentive offerings strategically timed work breaks and caused artificial price surges (Möhlmann & Henfridsson, 2019), circumventing AI's intended organizational purpose. Second, when the AI is superior to humans, relinquishing control to an inferior human creates what automation experts refer to as a prothesis problem (Muir, 1987). Muir (1987) pointed out that humans become unqualified supervisors of technology and more effective at tasks than humans. She asked how a person is supposed to evaluate or supervise a superior machine accurately. A third problem that does not affect the potency of the AI's application but could affect trust is that developers have given people control of things that ultimately do not matter to the function of the AI, like design features (Burton et al., 2020). Using control just as a psychological gimmick to achieve a level of trust could conceivably backfire. Therefore, discovering underlying reasons why augmentation elicits positive feelings (i.e., benevolence perceptions) may provide an alternative strategy to increase affective trust while avoiding the pitfalls of relinquishing control.

Decision Refinement as Benevolence Signal

In Logg et al.'s (2019) study where people overwhelmingly used algorithms to help them, it is important to note that the usage of the algorithm was markedly different in their study versus the algorithm aversion studies. Logg et al. (2019) used the Judge Advisor System paradigm to "measure the extent to which people assimilate advice" (p. 92) from the algorithms for tasks with which the person had no expertise. Appreciation was determined by the amount participants that changed their predictions based on algorithm suggestions. I believe that the stark difference in Logg et al.'s (2019) research compared to Dietvorst et al.'s (2015) aversion research (where participants had to choose to use a human advisor or an algorithm) stemmed from Logg et al.'s more benevolent role for the AI where the participant was given the AI specifically for their general welfare (to refine predictions to help achieve a prize). However, Logg et al. (2019) did not measure how participants felt about the AI.

Understanding if benevolence perceptions underpin the lure of augmented AI could shed light on the AI evaluative process. Moreover, it begs the question of whether just describing an AI as augmenting a worker is an effective intervention to improve affective trust. My theoretical argument is that benevolence is something that can be signaled (like by emphasizing augmentation), which is associated with increased affective trust. Because the context of AI for human augmentation may in and of itself signal benevolence, I hypothesize that the increased amount of benevolence perceived from AI described as augmenting a worker (versus not described as augmenting a worker) will partially mediate the positive relationship between AI and affective trust, which has been born-out through AI usage research. Thus, I posit:

H2: *Describing AI as having an augmenting role for a worker (versus not described with an augmenting role) will be positively associated with affective trust.*

H3: *Augmented role will be positively associated with perceived benevolence.*

H4: *Perceived benevolence will partially mediate the positive effect of the algorithm's augmented purpose on affective trust.*

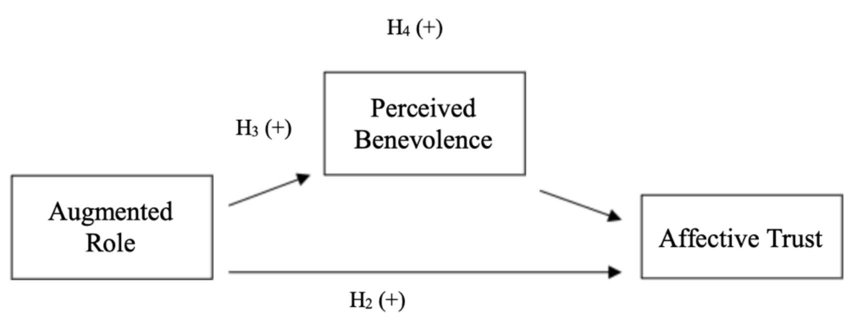


Figure 3. Perceived benevolence's mediating effects, partially explaining augmented role's positive association with affective trust.

Utility of Benevolence Messaging

There is little research validating practical interventions for marketers and managers to use to improve how workers or customers react to AI applications (Castelo et al., 2019). Describing AI as augmenting a worker may be one way that benevolence could be signaled. Another way to do so without evoking augmentation is by describing the benevolent intention of the AI. If low affective trust in AI applications could be improved by benevolent messages about AI intention, this would mean that benevolent perceptions are malleable. Thus, benevolence signaling (via messages) could become a valuable strategy to help reach the trust threshold needed to trigger use—especially early use.

Based on Schepman and Rodway's (2020) general negative attitudes surveying, we know that when AI is perceived as not benevolent, or having benevolence violations such as worker replacement or death from error, we expect affective trust in those applications to be lower. A good test of benevolence messaging is to test if it can modify comfort levels in AI associated with perceptions of likelihood of worker replacement and likelihood of death from an error.

Worker Replacement Concern

Scholars note that people are sensitive to AI replacing human workers (Castelo et al., 2019; Huang & Rust, 2018; Schepman & Rodway, 2020). AI scholars estimate that by 2040, AI capabilities may reach human intelligence equivalence (Ferràs-Hernández, 2018). Algorithms described in a way suggestive of replacing people likely are viewed with low or no benevolence and as such stir discomfort, or lower affective trust. Thus, I posit:

H5: *Likelihood of worker replacement by AI will be negatively associated to affective trust in AI.*

Magnitude of Error (Death) Concern

Another concern that affects trust in algorithms is the fact that algorithms make errors and people are sensitive to technology errors (Dietvorst et al., 2015), especially when errors could mean death (Schepman & Rodway, 2020). Lee and Moray (1992) found that automation trust was proportional to the magnitude of the error more so than frequency. Magnitude of error refers to how far off the error is and how consequential the error could be. Algorithm attitude research has overlooked the effect of worst-case scenario considerations. Logg et al. (2019) found algorithm appreciation in a context where the worst-case scenario was not winning an entry into a $10 raffle. Although the superior performance of AI is heralded, the effect of what is the worst outcome of error is often unstated, leaving people to imagine how wrong the algorithm could be and the potential life or death consequences of the error.

Decision theorists have long argued that errors trigger worst-case scenario estimations (Xue, 2020). The magnitude of worst-case scenarios affect decisions (Gilboa & Schmeidler, 2004; Xue, 2020). Death is certainly a worst-case consideration. For example, it is imaginable that people want AI to be 100% accurate in landing a plane but accept much less accuracy from AI in identifying additional items to purchase from an online website. Emphasizing the importance of worst-case scenario consideration is the condition that when people use AI, such as automation, they can only deal with errors *after they happen* (Hoff & Bashir, 2015). As concerns increase about the worse-case outcome of an error, it is logical that comfort with that AI application will be less. Thus, I posit:

H6: *Likelihood of death from an AI error will be negatively associated with affective trust.*

The literature review has established that it is possible to evoke cognitive trust with cognitive messages. Therefore, I expect it will be possible to evoke affective trust with affective messages. Specifically, by comparing the effects of two descriptions of exactly the same AI applications—one with and one without benevolent messaging—I expect that benevolent messaging will increase affective trust even in AI applications known to be rated with very low comfort. It is logical to predict that negative feelings from low-benevolent AI (higher likelihood to replace a worker and higher likelihood of death from an error) will be softened by benevolent messaging. This intervention is meant to test whether benevolence messaging can quell the impact of low benevolence perceptions on affective trust. Thus, I posit:

H7&H8: *Benevolence messaging will moderate the negative association between likelihood of worker replacement and likelihood of death from error and affective trust such that this relationship will be weaker for AI described with benevolence messaging (versus without).*

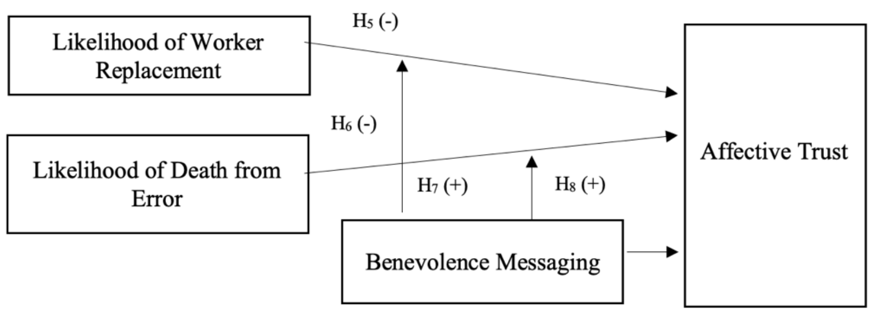


Figure 4. Perceived benevolence's moderating effects on perceptions of likelihood of non-benevolent outcomes (worker replacement and death from error).

In summary, affective trust in AI is an overlooked trust source. Affective trust's unresponsiveness to cognitive trust manipulations is contrary to the trust theory and thus has heightened interest in other affective trust antecedents, particularly the human affective trust base benevolence. Benevolence evaluation of AI could help explain affective trust in AI applications and explain current preferences for augmented AI. Benevolence messaging, explaining how the AI is intended for our general welfare, may help overcome negative attitudes about its use and boost affective trust, a type of AI trust that has been difficult to budge. Therefore, the present research helps to expand the understanding of AI evaluations, extends humanization as a way to build trust in AI and contributes to the human-AI trust theory. Ultimately, understanding the effects of benevolence perceptions could help managers and marketers promote AI and overcome delays in acceptance and adoption.

CHAPTER III

METHODS AND RESULTS

I conducted the following series of online surveys to explore the importance of a novel trustworthiness factor in AI attitude research—perceived benevolence of artificial intelligence. Three studies were conducted: (1) benevolence's relationship to affective trust (Hypothesis 1, Studies 1a and 1b), (2) the effect of manipulating AI descriptions with augmented intelligence descriptions and the role of benevolence to explain the difference in affective trust (Hypotheses 2, 3 and 4, Study 2), and (3) whether messages conveying the benevolent intent of AI could moderate the negative relationship between concerns with AI (worker replacement, death from error) and affective trust (Hypotheses 5, 6, 7 and 8, Study 3). All studies were approved by the Oklahoma State University Institutional Review Board and granted exempt status.

Study 1a Methods

The primary purpose of Study 1a was to gain an initial understanding of benevolence variance and relationship to affective trust. The secondary goal was to generate data for use in selecting AI applications for Studies 1b, 2 and 3.

Participants

In June 2022, I administered an online survey to 115 people recruited through Amazon's MTurk. The participants were paid \$3.75 each for completing the survey. The average age of participants was 38 years old (*SD = 9.84*) with the youngest being 20 and the oldest 67. Participants were female (31%) and male (64%); 1% chose "prefer not to say" and 4% did not respond. Seventy-eight percent (78%) were Caucasian, 8% Asian, 6% African American, 2% Hispanic, 1% Native American, 1% chose "other" and 4% did not respond. Regarding employment status, 77% reported working full-time, 10% working part-time, 3% unemployed and looking for work, 2% retired, 1% student, 1% homemaker or stay-at-home parent, 2% chose "other," and 4% chose not to respond. Regarding income, 8% made less than $25,000, 25% made $25,000 to 49,999, 28% made $50,000 to $74,999, 19% made $75,000 to $99,999, 10% made 100,000 to 149,000, 5% made more than 150,000, 1% preferred not to say, and 4% chose not to answer.

Procedures

An online survey was created using 20 AI descriptions from Schepman and Rodway's (2020) AI attitude research. Their study asked participants to rate how comfortable they were with several AIs. I used Schepman and Rodway's study to identify the top 10 highest-rated and top 10 lowest-rated AIs. I removed and replaced two of the low comfort AIs because they were duplicates. For example, I removed "providing psychotherapy for patients with phobias" because it appeared to be the same as "providing psychological counseling." The duplicative AIs were replaced by the next AI in order. See Table 2 for the 20 AIs selected.

Because AI attitudes are influenced by cognitive trust (Castelo et al., 2019; Schepman & Rodway, 2020) and may be influenced by familiarity (Belanche et al., 2019; Castelo et al., 2019; Logg et al., 2019), I controlled for cognitive trust and familiarity (see Figure 1) to test the unique relationship that benevolence has to affective trust. Therefore, the survey asked participants to rate benevolence, cognitive trust, familiarity, and affective trust. To limit common method variance, I gave participants confidentiality assurances and randomly presented the 20 applications and the four ratings for each AI (Podsakoff, MacKenzie, Lee & Podsakoff, 2003).

Measures

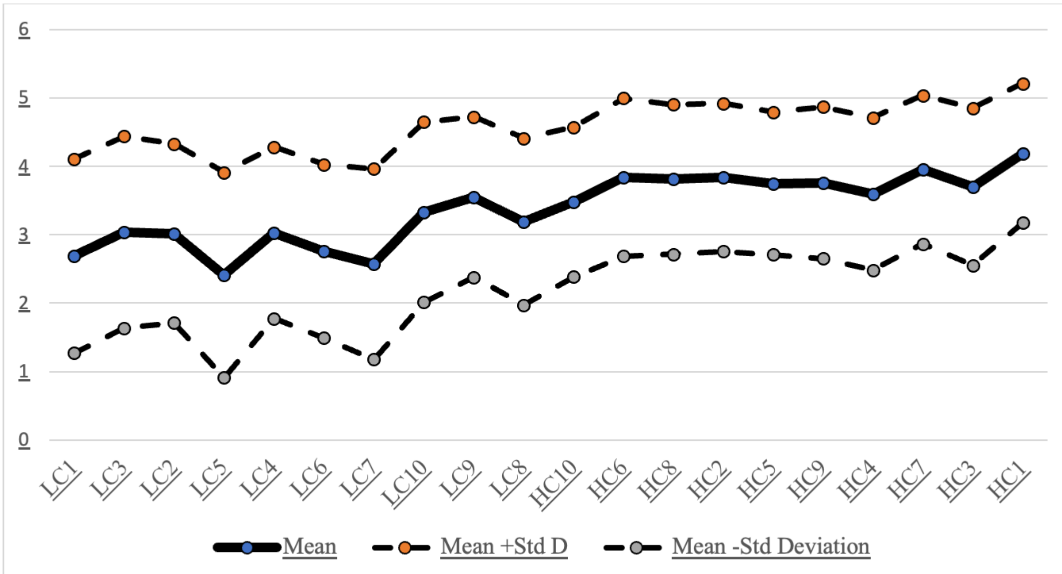
Due to the large number of AI applications to be rated by participants (20 AIs), I measured all variables (benevolence, affective trust, familiarity, and cognitive trust) with one-item measures. Length of survey and numerous referents are contexts where single item surveys can increase survey attentiveness and participation (Matthews, Pineault & Hong, 2022). Because single items directly reflect the construct definition, they tend to increase construct validity (Matthews et al., 2022). Additionally, using one-item measures to assess AI attitudes follows other AI attitude research studies (Bonezzi & Ostinelli, 2021; Castelo et al. 2019; Schepman & Rodway, 2020).

Benevolence was measured by "Please rate to what extent using AI for this task is in humans' general welfare." Affective trust was measured by "Please rate how comfortable you are with AI doing this task." Cognitive trust was measured by "Please rate how capable AI is to do this task." Familiarity was measured by "Please rate how familiar you are with AI doing this task." Participants were asked to provide their responses to all four ratings using a 5-point Likert-type scale (1 = Not at all to 5 = Extremely).

Study 1a Results

The data were analyzed using IBM SPSS Statistics 27. The analysis was done at the individual level for each AI. Missing data were addressed using listwise deletion. Note that the high/low comfort categorizations of AI applications from Schepman and Rodway (2020) were replicated in this study although the precise ranking differed slightly; all 10 high comfort AIs were highest in affective trust in this present study and all 10 low comfort AIs were the 10 lowest in affective trust in this current study. The means, standard deviations, and intercorrelations are presented in Table 3 and the 20 regressions in Table 4.

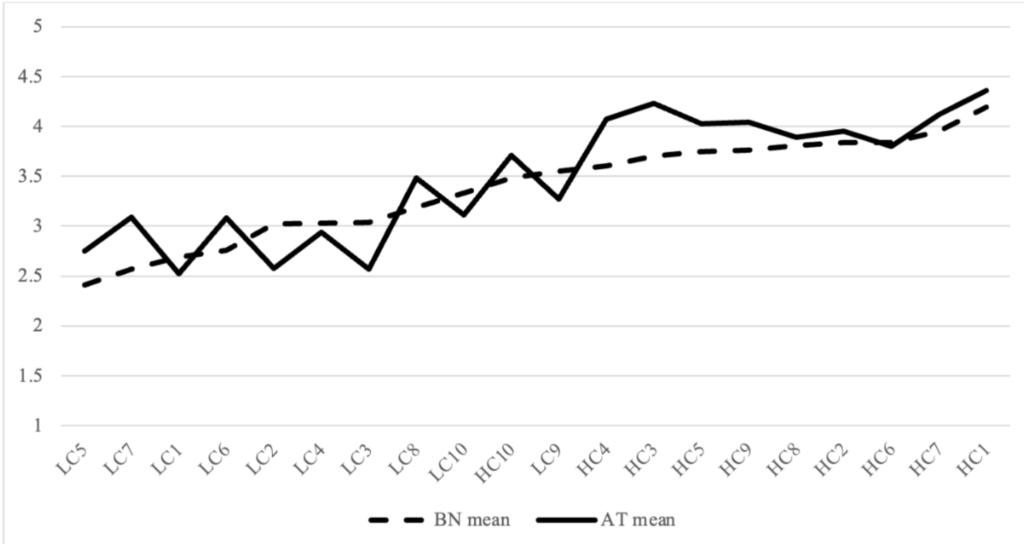
Since a goal of the study was to understand perceived benevolence's variance, benevolence's means and standard deviations were analyzed. Overall, benevolence means varied from highest "Translating speech into different languages" (M = 4.19, SD = 1.02) to lowest "Playing a team football match" (M = 2.41, SD = 1.50). Across all 20 applications, the average benevolence score was M = 3.38 (average SD = 1.21), confirming that individuals' perceived benevolence of AI applications varies between individuals and across AIs. Graph 1 presents a visual for how benevolence means and +/- standard deviations varied among the 20 AIs.



LC = Low Comfort
HC = High Comfort

Graph 1. Benevolence means and standard deviations from lower to higher affective trust means.

Exploring this relationship further, Graph 2 presents affective trust and benevolence means from lower to higher benevolence means. This provides evidence for the premise of my dissertation that benevolence perceptions are positively associated with affective trust.



HC= High Comfort
LC= Low Comfort

Graph 2. Benevolence and affective trust means from lower to higher benevolence means.

Further supporting the association between benevolence and affective trust, all 20 bivariate correlations between the two are statistically significant at $p < 0.05$ (see Table 3). The average correlation between benevolence and affective trust is $r = 0.60$. The highest correlation is for "Predicting relationship breakdowns via listening smart home devices" ($r = .78$, $p < 0.01$), and the lowest correlation is for "Translating speech into different languages in real-time" ($r = .39$, $p < 0.01$). These results support this study's theoretical argument that similar to interpersonal relationships, benevolence perceptions of AI are associated with affective trust in AI.

Hypothesis 1 Test

Hypothesis 1 proposed that benevolence and affective trust would be significantly associated after controlling for cognitive trust and familiarity. Because cognitive trust, familiarity and benevolence are related, I conducted hierarchical linear regression to test if benevolence had a unique and independent effect on affective trust beyond cognitive trust and familiarity. In the first step, I first regressed familiarity and cognitive trust on affective trust. For the second step, I regressed benevolence on affective trust for each AI. Results are presented in Table 4.

Benevolence significantly predicted affective trust in Step 2 in 17 of the 20 (85%) models; in all 17, it had a positive, direct effect on affective trust. The largest significant RSquare change was 0.22 ("Forecasting storm damage in forestry plantations") and the smallest significant change was 0.01 ("Selecting staff for employment"). Cognitive trust was significantly, positively associated with affective trust in 19 of 20 (95%) models while familiarity was significantly related in only six of 20 models (30%). I concluded that because benevolence had a unique, significant, and positive effect on affective trust above and beyond the controls in almost all models, Hypothesis 1 is mostly supported. Reasons as to why benevolence was not significant in certain models could be explored in future research.

Given familiarity's infrequent significant association to affective trust in regression, I further explored its association with affective trust and its unique effect. Sixteen of 20 (80%) familiarity-affective trust bivariate correlations were statistically significant. Yet, when I investigated the unique effect of familiarity using hierarchical regression (by regressing first cognitive trust and benevolence on affective trust and secondarily adding familiarity), familiarity was statistically significant in only one of 20 (5%) models ("Predicting relationship breakdowns via listening smart home devices") and changed the significance of benevolence in only one of 20 (5%) models ("Playing a team football match"). Familiarity across AI applications was generally low (average $M = 2.65$, $SD = 1.07$). The familiarity mean exceeded $M = 3.0$ (3 = Somewhat) for only four of 20 (20%) AIs ("Translating speech into different languages in real time," "Working in car manufacturing plants," "Selecting staff for employment," and "Driving a car") and in only one of those ("Selecting staff for employment") was familiarity significant. Interestingly, in six of 20 (30%) AI applications, the coefficient for familiarity was negative; in one case, the negative coefficient was statistically significant ("Helping farmers remove weeds and collect the harvest"). Like other research reporting mixed effects with familiarity (Berger et al., 2021), my study also found mixed results with affective trust. In some cases, the relationship was positive and in other cases it was negative.

Study 1b Methods

The primary purpose of Study 1b was to further test the association between benevolence and affective trust while controlling for cognitive trust and familiarity by using multi-item measures instead of single-item measures as used in Study 1a. Because the multi-item measures were initially written to measure interpersonal relationships, the secondary goal of Study 1b was to adapt cognitive trust, affective trust, and benevolence measures to the AI context and validate the hypothesized four-factor model.

Participants

In August of 2022, I administered an online survey to 161 people recruited through Amazon's MTurk. The participants were paid $3.75 each for completing the survey. The average age of participants was 36 years old (SD = 0.81), with the youngest being 22 and the oldest 63. Participants were 39% female, 61% male. Seventy-eight percent (78%) were Caucasian, 9% African American, 7.5% Asian, 2.5% Hispanic, 1% Native American, 1% Pacific Islander, 1% chose "other." Regarding employment status, 86% reported working full-time, 8% worked part-time, 1% were unemployed and looking for work, 1% student, 1% homemaker or stay-at-home parent, 3% chose "other."

Procedures

I created a survey with multi-item scales for each of the four constructs to increase measurement reliability. To keep survey length reasonable, I chose four AIs from Study 1a by first eliminating the three AI items where benevolence was not significantly related to affective trust. I then ranked each AI from one to 17 on three factors: the size of the RSquare change when adding benevolence to the model, the size of the standardized benevolence coefficient, and the affective trust mean score. I chose the RSquare change criterion because I wanted AIs with more distinction between benevolence and cognitive trust for discriminant validity, and the size of the benevolence coefficient to maximize variance explained by benevolence. I focused on the affective trust mean score because I was interested in understanding perceived benevolence's role in high affective trust. The three rankings were totaled to calculate a total score and I used the top four-ranked AIs. See scoring and rankings in Table 5.

The AIs were presented to participants in random order. The four constructs appeared in random order, and the order of the items for each construct appeared randomly but as a set. To further decrease common method variance, I reassured participants that their participation was anonymous. I did not collect any identifying information, and I instructed participants that there were no wrong answers (Podsakoff et al., 2003).

Measures

All responses were provided on a 5-point Likert scale (1 = strongly disagree to 5= strongly agree). Construct scales are included in Appendix B. Benevolence was measured by adapting Mayer and Davis' (1999) interpersonal items for an AI context, similar to how Kim, Kim, Lyons, and Nam (2020) adapted benevolence for robot attitude research. I added two items that focused on the perception that AI is in humans' welfare, corresponding with benevolence's definition. "AI doing this task is good for humans." "AI doing this task is in humans' best interest." There were seven benevolence items.

Cognitive trust was measured by adapting Mayer and Davis's (1999) interpersonal items for an AI context, similar to how Kim et al. (2020) adapted cognitive trust for robot attitude research. In a series of pilot studies, the adapted item "I think AI has much knowledge about doing this task" was dropped for having low item-total correlations. Upon item content review, the interpersonal trait of "having knowledge" did not fit with AI conceptually; therefore, that item was dropped. There were five cognitive trust items. Familiarity was measured by adapting Gefen's (2000) scale for familiarity with Amazon.com to familiarity with AI. There were three familiarity items.

Affective trust was measured by adapting Mayer and Davis' (1999) interpersonal scale and followed Lyon and Guznov's (2019) adaption for a robot. In pilot testing, the reverse-worded item "I would feel comfortable with the AI if a person had a good way to monitor it" yielded low item-total correlations, as is typical for reverse-worded items (Tay & Jebb, 2018). Therefore, the reverse-coded item was dropped. I added three items to the affective trust scale to better capture the construct within the context of AI. “AI that can perform this task is exciting.” “I feel positive about AI doing this task.” “I am optimistic about AI being used for doing this task.” There were six affective trust items.

Study 1b Results

I analyzed the data using IBM SPSS Statistics 27. The analysis was done at the individual level for each AI. Missing data were addressed using listwise deletion. The means, standard deviations, intercorrelations and Cronbach alphas are presented in Table 6.

Confirmatory Factor Analysis

To test that each construct was distinct, I conducted confirmatory factor analyses in MPlus comparing one, three and four-factor models. Given that affective trust and cognitive trust were highly correlated (Table 6), I compared a three-factor model with trust variables combined. In all four AI’s, the baseline four-factor model had the best fit to the data (Table 7). In all 4 four-factor models, RMSEA, CFI and SRMR indicated a reasonable approximate model fit to the data. Overall, confirmatory factor analyses results indicate that the four variables were distinct and that the four-factor model was theoretically sound.

Hypothesis 1 Test

Hypothesis 1 proposed that benevolence and affective trust would be significantly and positively associated in a model where cognitive trust and familiarity are controlled. To replicate the results from Study 1a, I conducted hierarchical regression, adding benevolence in the second step. Results are presented in Table 8. In all four AIs, benevolence had a significant, positive effect on affective trust. Consistent with Study 1a, cognitive trust was also positively related to affective trust. Also consistent with Study 1a, the relationship between familiarity and affective trust was inconsistent; significant in only two of the four AIs and negatively related to affective trust in three of the four AIs.

Study 2 Methods

In both practice and literature, there is an assumption that people trust AI more when it is described as augmenting (or collaborating) with humans, referred to as augmented intelligence. Therefore, Study 2 tested the relationship between framing AI as augmented intelligence (versus artificial intelligence) and affective trust, and also the mediating power of perceived benevolence to explain the difference, while controlling for the covariates of cognitive trust and familiarity.

Participants

In September of 2022, I administered an online survey to 203 people recruited through Amazon's MTurk; 100 were in one group and 103 were in another. The participants were paid $1.50 each for completing the survey. The average age of participants was 38 years old (*SD* = 10.40) ranging from 20 to 68. Forty percent (40%) were female, 52% male, 1% non-binary, and 7% chose not to answer. Participants' race was 81% Caucasian, 4% African American, 3% Asian, 3% Hispanic, 1% Native American, 1% chose "other," and 7% chose not to answer. Regarding employment status, 84% reported working full-time, 5% worked part-time, 1% were unemployed and looking for work, 1% retired, 1% student, 1% homemaker or stay-at-home parent, 1% chose "other," and the remaining chose not to answer.

Procedures

For Study 2, I selected AIs from the 17 with significant associations between benevolence and affective trust from Study 1a. I chose the 10 with the lowest affective trust scores based on the logic that describing AI as augmenting is usually done with AIs with which people are uncomfortable; therefore, the greatest and most valuable effect of an "augmentation" manipulation would be found among low affective trust AIs. I created two types of AI descriptions (with and without augmented intelligence description manipulation). Manipulations are included in Table 9. I used the AI descriptions from Study 1a for the artificial intelligence descriptions and used the same 10 AIs to frame them as augmented intelligence by adapting the descriptions from Raisch and Krawkowski's (2021) study. Augmented intelligence is defined as, “augmentation means that humans collaborate closely with machines to perform a task” (Raisch & Krawkowski, 2021, p. 2). An example manipulation is “Artificial intelligence that selects staff for employment” that was reframed as “augmented intelligence that collaborates with recruiters to select staff for employment.” Participants were randomly assigned to one of the two groups (0= artificial intelligence description, 1=augmented intelligence manipulation) in which all the AI descriptions were either all artificial intelligence or augmented intelligence descriptions. In addition to the manipulation, I provided tailored instructions for each group including a statement of what type of AI they were rating (artificial intelligence or augmented intelligence) with a description of artificial intelligence as autonomous and a description of augmented intelligence as collaborative. Again, to decrease common method variance, participants were told their answers were anonymous. I did not collect any identifying information, and I used random assignment to conditions (Podsakoff et al., 2003).

Measures

All factors were measured with single-item measures given that the results of Study 1b (multi-item measures) replicated the results of Study 1a, which used the single-item measures. In addition, single-item measures correspond highly to construct definitions (Matthews et al., 2022). All responses were provided on a 5-point Likert scale. Benevolence was measured with “Please rate to what extent using AI for this task is in human’s general welfare” (1 = Not at all to 5 = Completely). Affective trust was measured with “Please rate how comfortable you are with AI doing this task;” familiarity with “Please rate how familiar you are with AI doing this task,” and cognitive trust with “please rate how capable AI is to do this task” (1 = Not at all to 5 = Extremely).

Study 2 Results

The regression analysis was completed with IBM SPSS Statistics 27, using Hayes’ PROCESS v4.1, Model 4 with 5,000 bootstraps. I conducted the analysis for each of the 10 pairs of AI descriptions (Group 0 = Artificial Intelligence, Group 1 = Augmented Intelligence). Table

10 presents the means, standard deviations and intercorrelations. Given that Study 1a and Study 1b demonstrated cognitive trust and familiarity were associated with affective trust, they were included as covariates. Table 11 presents the mediation regression results.

Hypotheses Tests

The differential framing of augmented intelligence and artificial intelligence was not related to affective trust in any of the 10 descriptions. Additionally, there were no significant effects of framing on perceived benevolence. Thus, there was no significant indirect effect from the groups through benevolence to affective trust. Therefore, H2, H3, and H4 were not supported. Although the framing manipulation was not associated with perceived benevolence and affective trust, note that the proposed model for affective trust in AI that was researched in Study 1a and Study 1b was again largely upheld—even among AI with the lowest affective trust scores. Affective trust was significantly associated with cognitive trust, benevolence and familiarity.

Study 3 Methods

The theory that people are concerned about AI replacing workers and/or AI being implemented in life and death scenarios has been used to explain low affective trust in AI and AI aversion. How to address these concerns has not been identified. Therefore, Study 3 tested the relationship between people's perception of AI applications replacing workers and safety with affective trust. I then tested the moderating power of perceived benevolence to lessen these likelihood perceptions' relationship to affective trust, controlling for cognitive trust and familiarity.

Participants

In September of 2022, I administered an online survey to 199 people recruited through Amazon's MTurk. Ninety-eight were in Group 0 (received no explanation of the benevolent intent of AI) and 101 were in Group 1 (received an explanation of the benevolent intent of AI). The participants were paid \$1.75 each for completing the survey. The average age of participants was 38 years old (*SD* = 11.17), ranging from 22 to 70. Forty-three percent (43%) were female, 57% male, and no person selected non-binary. Participants' race was 83% Caucasian, 6% Asian, 5% African American, 4% Hispanic, 1% Native American, and 1% chose "other." Regarding employment status, 89% reported working full-time, 6% worked part-time, 1% were unemployed and looking for work, 1% retired, 1% student, 1% homemaker or stay-at-home parent, 1% chose "other," and one chose not to answer.

Procedures

For Study 3, I selected five AI applications from Study 1a that were lowest in affective trust and could reasonably trigger workplace replacement and/or death from error concerns. For example, acting as a primary care doctor likely triggers the notion of AI replacing a doctor and an error could be associated with death. To ensure that only factual benevolent messages were used, I only used benevolent reasons for AIs development as documented in news articles, academic journals, or online reports. To operationalize benevolence, all benevolent intent messages included a statement that the AI was intended to improve human welfare by doing something to improve a problematic current state of an issue. An example of a benevolent message for driving a car is "This AI is intended to promote human welfare by reducing human errors which currently cause 94% of traffic deaths."

Participants were randomly assigned to a survey where either all AI descriptions had or did not have benevolent messaging included. Other than the benevolent message added to the manipulation group, the two descriptions were identical. The AI descriptions appeared in random order. Next, I administered the likelihood evaluations, benevolence and affective trust followed by controls cognitive trust, and familiarity, in that order. To decrease common method variance, participants were told their answers were anonymous. No identifying information was collected, and I used random assignment to conditions (Podsakoff et al., 2003).

Measures

All factors were measured with single-item measures. All responses were provided on a 5-point Likert scale. The likelihood of worker replacement was measured by "How likely is it that implementation of this AI would result in replacing workers?" The likelihood of death from error was measured by "How likely is it that this AI making an error would result in death?" (1 = Definitely Unlikely to 5 = Definitely Likely). Benevolence was measured with "Please rate to what extent using AI for this task is in human's general welfare" (1 = Not at all to 5 = Completely). Affective trust was measured with "Please rate how comfortable you are with AI doing this task;" familiarity with "Please rate how familiar you are with AI doing this task," and cognitive trust with "please rate how capable AI is to do this task" (1 = Not at all to 5 = Completely).

Study 3 Results

The moderation regression analysis was completed with IBM SPSS Statistics 27, utilizing centered likelihood variables interacting with the framing manipulation. I conducted the analysis for each of the five pairs of AI descriptions (Group 0 = no benevolent intent message, Group 1 = benevolent intent message). Table 13 presents the means, standard deviations and intercorrelations. In addition to the framing manipulation, I also measured benevolence; therefore, both benevolence and group are in the intercorrelation table. Table 14 presents the moderation regression results.

Hypotheses Tests

None of the five framing manipulations resulted in a significant difference nor did their interaction with the likelihood variables result in significant interactions. Interestingly, neither likelihood of worker replacement nor likelihood of death by error were statistically significantly associated with affective trust in the model. Therefore, H5, H6, H7, and H8 were not supported.

The unstandardized (not statistically significant) effects for likelihood of worker replacement and likelihood of death from error were low, ranging from 0.00 (*SE* = 0.08) to 0.10 (*SE* = 0.08) for likelihood of worker replacement and ranging from - 0.01 (*SE*= 0.06) to -0.19 (*SE* = 0.10). These results suggest that these concerns may not be as closely linked to affective trust levels as the theory has proposed.

I conducted a supplemental moderation analysis, using centered benevolence scores (instead of the framing manipulation) to generate the interaction with centered likelihood variables. The results are presented in Table 15. By using benevolence, the likelihood of death by error was significantly negatively associated with AI that drives a car (B = -0.18, SE = 0.06). Likelihood of worker replacement was significantly associated with affective trust for AI that acts as bank branch employee (B = 0.16, SE = 0.07). Note that this is a significant, positive effect (instead of negative as hypothesized), possibly explaining that people prefer not to deal with people for banking. Two interactions of the likelihood of death and benevolence were statistically significant.

In the supplemental analysis, I highlighted that the model for benevolence, familiarity, and trust was upheld in three of the five models, even when controlling for likelihoods of worker replacement and death from error and their interactions with benevolence, even among these AIs associated with the lowest affective trust means. This further supports the unique, significant effect of the perceived benevolence of AI on affective trust.

CHAPTER IV

CONCLUSION

Discussion

This dissertation study aimed to test a novel model for trust in AI, focusing on affective trust. AI research previously demonstrated that people's beliefs in the capability of AI (cognitive trust) differs from their comfort with AI (affective trust). Research also has determined that people's discomfort (low affective trust) with AI was a root cause for reluctance to use AI. For example, Dietvorst et al. (2015) found that people chose not to use AI even after personally witnessing it outperform a qualified human. Even though affective trust appears to be a critical barrier, researchers have almost exclusively focused on building cognitive trust with minimal attention to understanding affective trust.

The dual affective-cognitive trust dynamic found in AI research is supported by a leading interpersonal trust theory and research stream. According to McAllister (1995), affective trust is based on a perception that a person considers your welfare when they act. Mayer et al.'s (1995) interpersonal trust theory also identified that a perception of another's interest in a person's welfare was a basis of trust; they call it trust base benevolence. Therefore, I questioned if perceived benevolence of AI could help explain human-AI affective trust. Indeed, in all four studies, perceived benevolence and affective trust were consistently and significantly positively related (39/39 intercorrelation tests).

In addition to perceived benevolence, literature points to two other consistent bases of affective trust: cognitive trust (capability perceptions) and familiarity. This four-factor model was supported by confirmatory factor analysis in four of four AI applications in Study 1b. The four-factor results give further credibility to the two distinct types of trust in AI. Across 24 tests in Study 1a and 1b, perceived benevolence was positively related to affective trust over and above cognitive trust and familiarity in 21 of 24 AIs. Hypothesis 1 was supported.

Perceived benevolence's contribution to explaining the affective trust variance is not slight. The average of the significant standardized coefficients from Study 1a for cognitive trust was 0.52 (average of 19 significant standardized coefficients, $B = 0.52$) and for perceived benevolence was 0.35 (average of 17 standardized coefficients, $B = 0.35$). The perceived benevolence magnitude of effect on affective trust was 70% of the amount of cognitive trust. Yet, researchers have overlooked the notion that people evaluate the benevolence of an AI application.

This dissertation also calls into question prior knowledge about affective trust. Scholars have encouraged developers and marketers to focus on AI in an augmenting role to gain better AI acceptance (Burton et al., 2020). Researchers theorized that augmentation increases comfort in AI by making it clear that people are "in the loop." The literature did not have direct comparisons of trust in AI versus trust in augmented intelligence to know if augmentation yields a higher level of trust. However, use studies suggest that AI that is collaborative is more accepted (e.g., the algorithmic advisor tool in studies by Logg et al., 2019). Therefore, I hypothesized that AI described as augmented intelligence with a collaborating role with a worker would echo use studies rendering higher affective trust.

Surprisingly, in a two-group study (Study 2) where one group rated 10 AI applications presented as autonomous "artificial intelligence" and another group rated the same 10 AI applications presented as "augmented intelligence" that "collaborates with" a specific worker, no significant difference in affective trust was found. Hypothesis 2 was not supported. There was also no increased perception of benevolence created by the augmentation description and no indirect effect through benevolence. Hypotheses 3 and 4 were not supported.

The failure of Hypothesis 2 (the lack of an "augmented intelligence" manipulation effect on affective trust) suggests that if augmentation increases affective trust, as researchers have claimed, it is not done at the time of initial trust. Notably, it was in actual use studies where people's positive reaction to augmenting AI was found. It is possible that improvement in benevolent perceptions of augmenting AI may also occur during and after use. Based on the results of the present study, individuals have similar initial perceptions of "augmented intelligence" as they do of "artificial intelligence," calling into question the fruitfulness of efforts in the field to position AI as only augmenting to improve initial trust and adoption. The augmentation manipulation's failure also supports Castelo et al.'s (2019) contention that trust is highly related to the task itself.

Directly comparing artificial intelligence descriptions and augmented intelligence descriptions was novel. The finding that descriptions of augmenting, collaborative AI did not increase affective trust was unexpected. It was counter to "common-sense" recommendations from scholars and field efforts trying to improve initial trust in AI. For this dissertation, the lack of difference in affective trust between the two AI descriptions rendered mute the need for benevolence to explain it. And the takeaway for benevolence is the discovery that benevolence of AI is not signaled to people by describing it as collaborating with a worker.

Finally, in a two-group study, I attempted to manipulate benevolence to test if it would moderate the negative effect of worker replacement and death from error concerns affecting AI attitudes according to attitude studies (e.g., Schepman & Rodway, 2020). The manipulation was intended to boost affective trust in AIs by describing how the AI is intended for human welfare. The benevolence manipulation did not work. Some of my research choices may have impacted the lack of statistically significant results. First, I chose AI applications with low affective trust with the logic that they may be more affected by worker replacement or death from error scenarios and therefore, there was greater potential to boost affective trust. Due to other factors not yet understood, the applications rated lowest in affective trust may be more resistant to any efforts at improving trust in AI. It is unknown whether the manipulation would have been more effective with moderate or higher-rated AIs as people may be more open to learning about the human welfare intent of AIs with which they already have some comfort.

Second, I chose to operationalize the benevolent condition by simply directly informing participants that the AI was intended for human welfare and including a truthful statement defining the current problem negatively affecting humans. I did not explore different methods of framing human welfare gains (i.e., addressing worker shortage, alleviating human suffering, and timely or widespread access to solutions). Directly informing people about AI's human welfare intent may be too weak of a benevolence signal. When comparing means between the two manipulation groups, benevolence means were higher for three of the five applications described with benevolence messaging (AI for primary care doctor, bank branch employee, and prioritize aid); and, in all three of those cases, affective trust means were also higher. So, there is some evidence that perceived benevolence might be boostable with stronger benevolence signaling.

Although this dissertation did not identify how to signal benevolence, it found two ways that it does not. Given the significant association between affective trust and benevolence, it seems premature to conclude that benevolence messaging cannot be boosted and therefore cannot boost affective trust. What is needed is a better understanding of how benevolence perceptions are formed. What are the significant antecedents of benevolence of AI? I propose that question be resolved in future research to improve benevolence messaging.

Some of the most surprising findings of this dissertation were the relationships between the likelihoods of bad outcomes (likelihood of worker replacement and likelihood of death from error) and affective trust in the Study 3 model. The only significant relationship found was between likelihood of death from error and affective trust for AI that drives a car, and the benevolent framing did not moderate this relationship. Indeed, framing did not moderate any of the relationships involving the likelihood effect; hence Hypotheses 7 and 8 were not supported.

Although it is common for people to explain away low affective trust due to humans' desire to avoid worker replacement or limit AI in life-and-death situations, no research, to my knowledge, has tested the relationship between these likelihoods and affective trust while controlling for other factors known to influence affective trust. Study 3 did just that; seven of the eight relationships between the likelihoods and affective trust were not statistically significant. Hypotheses 5 and 6 were not supported. Individuals may be less sensitive to these likelihoods than assumed. Even more unexpectedly, all four coefficients for likelihood of worker replacement, albeit not statistically significant and small, were positive (rather than negative); whereas all coefficients for likelihood of death from error, three of four were not statistically significant but were negative, as expected. People may be more ambivalent than assumed about the likelihoods of worker replacement and death, so that affective trust formation is even more puzzling.

When analyzing Study 3 using the participants' benevolence ratings of the AIs instead of framing it as the moderator, benevolence was a significant predictor in three of the five models and significantly interacted with likelihood of death from error in two of five AIs. Therefore, perceptions of benevolence seem to matter to affective trust; it is clear that the Study 3 framing attempt did not adequately signal benevolence.

Contribution to Literature and Theory

Glikson and Woolley (2020) reviewed the existing research regarding "determinants of human trust in AI" (p. 627). They categorized research findings into two types of trust—cognitive and affective—and sorted their findings by types of AI researched: robotic, virtual (e.g., chat boxes), and embedded (e.g., algorithms). They identified a robust amount of cognitive trust factors for all AI types and a comparatively thin amount of affective trust factors for any AI type. Worse yet, the least number of studies and trust factors were identified for affective trust in embedded/algorithmic AI—arguably, the type of AI that is most common in workforces. Instead, measuring affective reactions to the presence of robots and chat boxes describes the bulk of affective trust in AI research.

More troubling, this affective trust knowledge gap exists. However, researchers are working to identify strategies to overcome aversion to algorithmic AI. Reducing AI reluctance has proven difficult to do, even when people are convinced of the AI's capability. How people feel about the AI was determined to be a barrier. Discomfort with AI (low affective trust) seems to be blunting cognitive trust gains. Yet, as demonstrated by Glikson and Woolley's (2020) review, researchers have largely ignored affective trust, especially in the case of algorithmic AI. This disconnect between the trust barriers impeding AI adoption (affective) and current trust in AI knowledge (cognitive) needs to be addressed. Therefore, I aimed to contribute knowledge regarding affective trust in AI.

Due to minimal human-AI affective trust research, I turned to human-human affective trust theory to find theoretical bases of affective trust. McAllister's (1995) trust theory contends that there are two distinct types of trust (cognitive-based and affective-based), which, combined, form trust. McAllister's theory fits with AI research asserting the same cognitive and affective trust dynamic. How cognitive trust and affective trust theoretically combine is interesting to AI usage barriers because cognitive trust is first, and then it flows to affective trust (Johnson & Grayson, 2005; Lewis & Weigert, 1985; McAllister, 1995; Schaubroeck et al., 2011). This cognitive trust → affective trust order theoretically supports AI trust researchers who expected cognitive trust to increase affective trust. However, researchers have discovered that increases in cognitive trust do not necessarily translate to affective trust increases, revealing that other bases of affective trust are at play.

McAllister (1995) described the affective trust base as a perception that the other party would act in a person's welfare. This same trust base is echoed in Mayer et al.'s (1995) competing interpersonal trust theory. Mayer et al. (1995) went on to empirically establish a relationship between benevolence and trust. Mayer and colleagues have argued ardently against the existence of affective trust, putting the two theories at odds with each other. However, when comparing the bases of trust in their two models, the two theories are in actuality very similar—the concept of benevolence overlap as a prime example. Therefore, I integrated benevolence from Mayer and colleagues' model into McAllister's model of affective trust. I departed from Mayer and colleagues' method by not modeling benevolence flowing to cognitive trust (the only type of trust Mayer et al. recognized) but rather to affective trust as McAllister posited. The resulting model was a novel trust in AI model with particular interest regarding whether people had varying perceptions of AIs' benevolence and if that was associated with the affective trust variance not accounted for by cognitive trust. Since familiarity was cited in both interpersonal and AI research as increasing trust, I included familiarity as a control variable.

By using the human-human trust theory, I discovered a potentially overlooked factor in AI trust research, perceived benevolence, as well as a theoretical explanation for how cognitive trust and affective trust relate to AI. Because cognitive trust flows to affective trust, there is a theoretical foundation for low affective trust blunting cognitive trust gains—the aversion phenomenon found in experiments and the field.

In sum, the theoretical model I posed overwhelmingly held, explaining affective trust in AI when that trust was either high or low. Although the focus of this dissertation was exploring and establishing the role of the perceived benevolence of AI and its relationship to affective trust, an overall trust model was also tested. The context where this theoretical model held was at the time of initial trust. Because initial trust in technology endures even long after use, the initial trust attitudes measured in this study were essential. My dissertation findings contribute to increased knowledge of affective trust in AI, which has been elusive to explain.

Practical Implications

Fostering trust in AI and avoiding aversion are challenges facing practitioners with few tools. In the absence of tools and in an effort to sideline aversion, marketers have reverted to avoiding mentioning that AI is utilized in products (Hengstler et al., 2016). I would argue that this is a precarious choice because when users learn later that AI was involved, the choice to hide the AI can backfire (Eslami, Rickman, Vaccaro et al., 2015), likely compounding trust problems. Therefore, understanding how to message information about AI could help managers and marketers to plan practical implementations of AI tools and get buy-in from humans using AI.

This dissertation did not establish that benevolence perceptions are malleable. Affective trust remained resistant to influence using the manipulations in the present study, but three valuable practical lessons were learned: (1) all the fuss about emphasizing augmented intelligence versus artificial intelligence may be for naught, or potentially only relevant among certain adoption targets; (2) simply stating that the AI is intended for human welfare is not enough to generate significantly higher benevolent perceptions; and (3) expecting people to warm-up to AI by becoming more familiar with it does not seem to be a reliable strategy to build trust. Nevertheless, from this study's findings, we now know that perceived benevolence of the AI matters to affective trust above and beyond cognitive trust and familiarity. Pursuing whether other benevolence framings are more effective seems worthwhile given that we currently do not have tools to improve affective trust. Understanding affective trust is salient to the future of work.

Limitations and Future Research

A limitation of the study is that I used a convenience sample recruited through MTurk. Therefore, it is uncertain if respondents took the survey seriously, limiting the generalizability of the survey findings. The majority of the respondents (84%) were full-time employees in workforces, which could help the generalizability of workforces. In Study 1a and 1b, the analysis was by individual, and all participants self-reported, so results are subject to common method variance. I took steps to limit common method bias in the first two studies by collecting no identifying information, randomization of measures, and reassurances that there were no wrong answers. In Study 2 and Study 3, the groups were randomly assigned to conditions, therefore, limiting common-method variance concerns.

Understanding the effects of benevolence may have been limited in two ways. First, I asked participants to what extent the AI is in humans' welfare, eliciting a general rating rather than asking to the degree that the AI is in their personal welfare, which may or may not be more compelling. In interpersonal trust at work, the perception of benevolence can be a general orientation [i.e., rating top leadership (Mayer & Davis, 1999)]; however, the welfare people are concerned about most often in human-human trust is their own, and those attitudes may or may not differ from general welfare attitudes. Second, I did not investigate if there was a within-person difference between AI described with and without benevolence cues. Within-person differences resulting from the same person being exposed to two messages may tell a different story than between-person. Within-person changes may be more relevant for practitioners who will likely be tasked with shifting customers' or employees' avoidant responses. These limitations could be addressed in future research.

I propose for future research to embrace cognitive trust and perceived benevolence as antecedents of affective trust in AI due to their durability across a variety of AIs. For the purposes of this dissertation, familiarity worked as a control, ruling it out as an unaccounted-for factor that strengthens cognitive trust and benevolence findings. Future research focused on the antecedents of perceived benevolence of AI may reveal effective ways to signal benevolence. Given how difficult influencing affective trust has been—combined with results that challenge the validity of what we think increases it (augmenting role, familiarity) or decreases it (likelihoods of worker replacement or death by error)—further research is warranted before we can confidently address the human reluctance burdening AI adoption. As for this dissertation, I took a step forward confirming the notion that people assess the extent to which AI is in humans' welfare (benevolence) and that it matters to affective trust.

REFERENCES

American Medical Association. (2022). *Augmented intelligence in medicine*. Retrieved January 10, 2022 from, https://www.ama-assn.org/practice-management/digital/augmented-intelligence-medicine

Bagaric, M., Hunter, D., & Stobbs, N. (2019). Erasing the bias against using artificial intelligence to predict future criminality: Algorithms are color blind and never tire. *University of Cincinnati Law Review*, *88*(4), 1037. https://scholarship.law.uc.edu/uclr/vol88/iss4/3

Belanche, D., Casaló, L. V., & Flavián, C. (2019). Artificial intelligence in FinTech: Understanding robo-advisors adoption among customers. *Industrial Management & Data Systems*, *119*(7), 1411–1430. https://doi.org/10.1108/IMDS-08-2018-0368

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. https://doi.org/10.1007/s12599-020-00678-5

Bonezzi, A., & Ostinelli, M. (2021). Can algorithms legitimize discrimination? *Journal of Experimental Psychology: Applied*, *27*. https://doi.org/10.1037/xap0000294

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

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

Chatterman. V., Kwon, W. -S., Gilbert, J. E., & Li, Y. (2014). Virtual shopping agents. *Journal of Research in Interactive Marketing*, *8*(2), 144-162.

de Visser, E., Monfort, S., Mckendrick, R., Smith, M., Mcknight, P., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, *22*. https://doi.org/10.1037/xap0000092

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

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. https://doi.org/10.1287/mnsc.2016.2643

Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Applied Psychology*, *88*(5), 736-748. https://psycnet.apa.org/doi/10.1037/0022-3514.88.5.736

Dzieżyk, M., Hetmańczuk, W., & Traczyk, J. (2020). Sensitivity to interventions and the relationship with numeracy. *Decyzje*, *2020*(34), 67–90. https://doi.org/10.7206/DEC.1733-0092.147

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 153–162. https://doi.org/10.1145/2702123.2702556

Fan, W., Liu, J., Zhu, S., & Pardalos, P. M. (2020). Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research*, *294*(1–2), 567–592. https://doi.org/10.1007/s10479-018-2818-y

Ferràs-Hernández, X. (2018). The future of management in a world of electronic brains. *Journal of Management Inquiry*, *27*(2), 260–263. https://doi.org/10.1177/1056492617724973

Gefen, D. (2000). E-commerce: The role of familiarity and trust. *Omega*, *28*(6), 725–737. https://doi.org/10.1016/S0305-0483(00)00021-9

Gilboa, I., & Schmeidler, D. (2004). Maxmin expected utility with non-unique prior. In *Uncertainty in economic theory* (pp. 141-151). Routledge.

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

Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, *105*, 105–120. https://doi.org/10.1016/j.techfore.2015.12.014

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Huang, M.-H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, *21*(2), 155–172. https://doi.org/10.1177/1094670517752459

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, *58*(4), 500–507. https://doi.org/10.1016/S0148-2963(03)00140-1

Johnson-George, C., & Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology*, *43*(6), 1306–1317. http://dx.doi.org/10.1037/0022-3514.43.6.1306

Jones, G. R., & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *The Academy of Management Review*, *23*(3), 531–546.

Jung, M. F., Lee, J. J., DePalma, N., Adalgeirsson, S. O., Hinds, P. J., & Breazeal, C. (2013). Engaging robots: Easing complex human-robot teamwork using backchanneling. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work - CSCW '13*, 1555. https://doi.org/10.1145/2441776.2441954

Kim, W., Kim, N., Lyons, J. B., & Nam, C. S. (2020). Factors affecting trust in high-vulnerability human-robot interaction contexts: A structural equation modelling approach. *Applied Ergonomics*, *85*, 103056. https://doi.org/10.1016/j.apergo.2020.103056

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, *35*(10), 1243–1270. https://doi.org/10.1080/00140139208967392

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces*, *63*(4), 967-985.

Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems, 17*(1), 39–71. https://doi.org/10.1016/j.jsis.2008.01.001

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. https://doi.org/10.1016/j.obhdp.2018.12.005

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

Lyons, J. B., & Guznov, S. Y. (2019). Individual differences in human–machine trust: A multi-study look at the perfect automation schema. *Theoretical Issues in Ergonomics Science*, *20*(4), 440–458. https://doi.org/10.1080/1463922X.2018.1491071

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Matthews, R. A., Pineault, L., & Hong, Y.-H. (2022). Normalizing the use of single-item measures: Validation of the single-item compendium for organizational psychology. *Journal of Business and Psychology*, *37*(4), 639–673. https://doi.org/10.1007/s10869-022-09813-3

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology, 84*(1), 123-136. https://doi.org/10.1037/0021-9010.84.1.123

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709-734. https://doi.org/10.2307/258792

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24-59.

McGrath, P., Desai, K., & Junquera, P. (2019). Resistance is futile: How corporate real estate companies can deploy artificial intelligence as a competitive advantage. *Corporate Real Estate Journal*, *9*(2), 121–129.

Mende, M., Scott, M. L., van Doorn, J., Grewal, D., & Shanks, I. (2019). Service robots rising: How humanoid robots influence service experiences and elicit compensatory consumer responses. *Journal of Marketing Research*, *56*(4), 535–556. https://doi.org/10.1177/0022243718822827

Möhlmann, M., & Henfridsson, O. (2019). What people hate about being managed by algorithms, according to a study of Uber drivers. *Harvard Business Review*. Retrieved from, https://hbr.org/2019/08/what-people-hate-about-being-managed-by-algorithms-according-to-a-study-of-uber-drivers

Möhlmann, M., Zalmanson, L., Henfridsson, O., & Gregory, R. W. (2021). Algorithmic management of work on online labor platforms: When matching meets control. *MIS Quarterly*, *45*(4), 1999–2022. https://doi.org/10.25300/MISQ/2021/15333

Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, *27*(5), 527–539. https://doi.org/10.1016/S0020-7373(87)80013-5

Musen, M. A., Middleton, B., & Greenes, R. A. (2014). Clinical decision-support systems. In E. H. Shortliffe & J. J. Cimino (Eds.), *Biomedical informatics* (pp. 643–674). Springer London. https://doi.org/10.1007/978-1-4471-4474-8_22

Obaid, M., Sandoval, E. B., Złotowski, J., Moltchanova, E., Basedow, C. A., & Bartneck, C. (2016). Stop! That is close enough. How body postures influence human-robot proximity. *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 354–361. https://doi.org/10.1109/ROMAN.2016.7745155

Ore, O., & Sposato, M. (2021). Opportunities and risks of artificial intelligence in recruitment and selection. *International Journal of Organizational Analysis* (ahead-of-print). https://doi.org/10.1108/IJOA-07-2020-2291

Pezzo, M. V., & Beckstead, J. W. (2020). Patients prefer artificial intelligence to a human provider, provided the AI is better than the human: A commentary on Longoni, Bonezzi and Morewedge (2019). *Judgment and Decision Making*, *15*(3), 443-445.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, *36*(6), 691–702. https://doi.org/10.1002/for.2464

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation-augmentation paradox. *Academy of Management Review*, 46(1). https://doi.org/10.5465/2018.0072

Reis, L., Maier, C., Mattke, J., Creutzenberg, M., & Weitzel, T. (2020). Addressing user resistance would have prevented a healthcare AI project failure. *MIS Quarterly Executive*, *19*(4), 279–236. https://doi.org/10.17705/2msqe.00038

Schaubroeck, J., Lam, S. S. K., & Peng, A. C. (2011). Cognition-based and affect-based trust as mediators of leader behavior influences on team performance. *Journal of Applied Psychology*, *96*(4), 863–871. https://doi.org/10.1037/a0022625

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports*, *1*, 100014. https://doi.org/10.1016/j.chbr.2020.100014

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *The Academy of Management Review*, *32*(2), 344–354. https://doi.org/10.2307/20159304

Tay, L., & Jebb, A. T. (2018). Establishing construct continua in construct validation: The process of continuum specification. *Advances in Methods and Practices in Psychological Science*, *1*(3), 375–388. https://doi.org/10.1177/2515245918775707

Wang, W., & Benbasat, I. (2007). Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems*, *23*(4), 217–246. https://doi.org/10.2753/MIS0742-1222230410

Xue, J. (2020). Preferences with changing ambiguity aversion. *Economic Theory*, *69*(1), 1–60. https://doi.org/10.1007/s00199-018-1156-2

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, *32*(4), 403–414. https://doi.org/10.1002/bdm.2118

Zhang, B., & Dafoe, A. (2019). Artificial intelligence: American attitudes and trends. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3312874

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. https://doi.org/10.1016/j.ijhcs.2016.12.008

APPENDIX A

TABLES

Table 2.

Study 1a AI selection from rankings by comfort from Schepman and Rodway (2020)

Top 10 by Comfort Ratings	Bottom 10 by Comfort Ratings
1. Translating speech into different languages in real time.	1. Predicting relationship breakdowns via listening smart home devices.
2. Teaching people sign language.	2. Providing psychological counseling.
3. Searching for life on other planets.	3. Being a primary care physician (Acting as a doctor in a general practice).
4. Working in car manufacturing plants.	4. Selecting staff for employment.
5. Forecasting storm damage in forestry plantations.	5. Playing a team football match.
6. Using smells in human breath to detect illness.	6. Providing Psychotherapy for patients with phobias*
7. Helping farmers remove weeds and collect the harvest.	7. Being a news anchor.
8. Discovering new chemical molecules for pharmaceutical or industrial applications.	8. Performing surgical procedures on patients*
9. Checking large volumes of documents for relevant legal evidence.	9. Being an actor in a film.
10. Reducing fraud related to exams or assessments.	10. Being a bank branch employee.
	11. Driving a car.
	12. Deciding how to prioritize aid during a humanitarian crisis.

**Removed from the list due to duplication.*

Table 3

Study 1a: Means, Standard Deviations, and Intercorrelations

		N	*M*	*SD*	1	2	3	4
HC1 Translating speech into different languages in real time	1. Familiarity	110	3.65	1.17	1			
	2. Cognitive Trust	110	4.25	0.89	0.16	1		
	3. Benevolence	110	4.19	1.02	0.18	0.48*	1	
	4. Affective Trust	110	4.36	0.84	0.21*	0.53*	0.39*	1
HC2 Teaching people sign language	1. Familiarity	110	2.42	1.51	1			
	2. Cognitive Trust	110	3.92	1.07	0.26*	1		
	3. Benevolence	111	3.84	1.08	0.19*	0.58*	1	
	4. Affective Trust	110	3.95	1.00	0.10	0.66*	0.62*	1
HC3 Searching for life on other planets	1. Familiarity	111	2.82	1.40	1			
	2. Cognitive Trust	111	3.90	1.03	0.34*	1		
	3. Benevolence	111	3.70	1.15	0.32*	0.33*	1	
	4. Affective Trust	111	4.23	0.90	0.02	0.42*	0.40*	1
HC4 Working in car manufacturing plants	1. Familiarity	110	3.44	1.24	1			
	2. Cognitive Trust	111	4.20	0.90	0.50*	1		
	3. Benevolence	111	3.60	1.11	0.52*	0.39*	1	
	4. Affective Trust	111	4.07	0.99	0.42*	0.57*	0.42*	1
HC5 Forecasting storm damage in forestry plantations	1. Familiarity	113	2.49	1.46	1			
	2. Cognitive Trust	113	3.81	0.93	0.28*	1		
	3. Benevolence	113	3.75	1.04	0.19*	0.42*	1	
	4. Affective Trust	113	4.03	1.02	0.08	0.45*	0.61*	1
HC6 Using smells in human breath to detect illness	1. Familiarity	110	2.16	1.49	1			
	2. Cognitive Trust	110	3.43	1.24	0.36*	1		
	3. Benevolence	110	3.84	1.15	0.08	0.53*	1	
	4. Affective Trust	110	3.80	1.20	0.27*	0.66*	0.66*	1
HC7 Helping farmers remove weeds and collect the harvest	1. Familiarity	111	2.62	1.54	1			
	2. Cognitive Trust	111	3.71	1.12	0.37*	1		
	3. Benevolence	111	3.95	1.09	0.24*	0.56*	1	
	4. Affective Trust	111	4.12	1.07	0.07	0.63*	0.57*	1
HC8 Discovering new chemical molecules for pharmaceutical or industrial applications	1. Familiarity	112	2.50	1.48	1			
	2. Cognitive Trust	112	3.83	1.05	0.22*	1		
	3. Benevolence	112	3.81	1.10	0.34*	0.58*	1	
	4. Affective Trust	112	3.89	1.13	0.26*	0.69*	0.57*	1
HC9 Checking large volumes of documents for relevant legal evidence	1. Familiarity	112	2.95	1.40	1			
	2. Cognitive Trust	112	3.96	1.04	0.28*	1		
	3. Benevolence	112	3.76	1.11	0.33*	0.46*	1	
	4. Affective Trust	112	4.04	1.05	0.41*	0.72*	0.61*	1
HC10 Reducing fraud related to exams or assessments	1. Familiarity	112	2.91	1.41	1			
	2. Cognitive Trust	113	3.84	0.96	0.30*	1		
	3. Benevolence	113	3.48	1.10	0.30*	0.44*	1	
	4. Affective Trust	113	3.71	1.14	0.33*	0.64*	0.54*	1
LC1 Predicting relationship breakdowns via listening smart home devices	1. Familiarity	111	2.45	1.52	1			
	2. Cognitive Trust	111	3.13	1.28	0.61*	1		
	3. Benevolence	111	2.69	1.42	0.62*	0.59*	1	
	4. Affective Trust	111	2.52	1.53	0.75*	0.60*	0.78*	1

Table 3 Continued.

		N	*M*	*SD*	1	2	3	4
LC2 Providing psychological counseling	1. Familiarity	111	2.46	1.41	1			
	2. Cognitive Trust	111	2.77	1.31	0.70*	1		
	3. Benevolence	111	3.02	1.31	0.58*	0.69*	1	
	4. Affective Trust	111	2.58	1.35	0.62*	0.75*	0.72*	1
LC3 Being a primary care physician (Acting as a doctor in a general practice)	1. Familiarity	110	2.18	1.34	1			
	2. Cognitive Trust	110	2.84	1.26	0.66*	1		
	3. Benevolence	110	3.04	1.40	0.56*	0.66*	1	
	4. Affective Trust	109	2.57	1.36	0.72*	0.74*	0.67*	1
LC4 Selecting staff for employment	1. Familiarity	111	3.08	1.32	1			
	2. Cognitive Trust	111	3.24	1.08	0.59*	1		
	3. Benevolence	111	3.03	1.25	0.51*	0.71*	1	
	4. Affective Trust	111	2.94	1.32	0.58*	0.77*	0.64*	1
LC5 Playing a team football match	1. Familiarity	110	2.14	1.48	1			
	2. Cognitive Trust	110	2.59	1.44	0.74*	1		
	3. Benevolence	110	2.41	1.50	0.90*	0.81*	1	
	4. Affective Trust	110	2.75	1.42	0.58*	0.67*	0.65*	1
LC6 Being a news anchor	1. Familiarity	111	2.28	1.42	1			
	2. Cognitive Trust	111	3.24	1.25	0.42*	1		
	3. Benevolence	111	2.76	1.27	0.54*	0.53*	1	
	4. Affective Trust	111	3.08	1.29	0.47*	0.65*	0.59*	1
LC7 Being an actor in a film	1. Familiarity	111	2.41	1.55	1			
	2. Cognitive Trust	111	2.80	1.41	0.71*	1		
	3. Benevolence	110	2.57	1.39	0.82*	0.71*	1	
	4. Affective Trust	111	3.09	1.40	0.54*	0.72*	0.58*	1
LC8 Being a bank branch employee	1. Familiarity	111	2.61	1.32	1			
	2. Cognitive Trust	111	3.58	1.01	0.43*	1		
	3. Benevolence	111	3.19	1.22	0.49*	0.43*	1	
	4. Affective Trust	111	3.48	1.21	0.38*	0.59*	0.58*	1
LC9 Driving a car	1. Familiarity	111	3.58	1.10	1			
	2. Cognitive Trust	111	3.49	1.04	0.44*	1		
	3. Benevolence	111	3.55	1.17	0.46*	0.60*	1	
	4. Affective Trust	111	3.27	1.27	0.39*	0.69*	0.70*	1
LC10 Deciding how to prioritize aid during a humanitarian crisis	1. Familiarity	114	2.39	1.53	1			
	2. Cognitive Trust	113	3.34	1.22	0.51*	1		
	3. Benevolence	114	3.33	1.32	0.44*	0.70*	1	
	4. Affective Trust	114	3.11	1.28	0.53*	0.80*	0.70*	1

$*p < 0.05$

Table 4.

Study 1a: Hierarchical Regression Analysis for DV Affective Trust

		Step 1		Step 2	
HC1 Translating speech into different languages in real time					
	N	*B*	*SE*	*B*	*SE*
Constant		1.98*	0.37	1.77*	.39
Familiarity	110	0.09	0.06	0.08	0.06
Cognitive Trust	110	0.48*	0.08	0.41*	0.09
Benevolence	110			0.13	0.08
R^2		0.30*		0.31*	
R^2 Change				0.02	
HC2 Teaching people sign language					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		1.58*	0.28	1.07*	0.28
Familiarity	110	-0.05	0.05	-0.06	0.05
Cognitive Trust	110	0.63*	0.07	0.44*	0.08
Benevolence	110			0.34*	0.08
R^2		0.43		0.52*	
R^2 Change				0.09*	
HC3 Searching for life on other planets					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		2.89*	0.31	2.33*	0.33
Familiarity	111	-0.09	0.06	-0.14*	0.06
Cognitive Trust	111	0.41	0.08	0.33*	0.08
Benevolence	111			0.27*	0.07
R^2		0.19		0.29*	
R^2 Change				0.10*	
HC4 Working in car manufacturing					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		1.38*	0.37	1.13*	0.38
Familiarity	110	0.15*	0.07	0.08	0.08
Cognitive Trust	111	0.52*	0.10	0.49*	0.10
Benevolence	111			0.18*	0.08
R^2		0.35		0.38*	
R^2 Change				0.03*	
HC5 Forecasting storm damage in forestry plantations					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		2.19*	0.37	1.21*	0.35
Familiarity	113	-0.03	0.06	-0.06	0.05
Cognitive Trust	113	0.50*	0.10	0.27*	0.09
Benevolence	113			0.51*	0.08
R^2		0.20		0.42*	
R^2 Change				0.22*	

Table 4. Continued

HC6 Using smells in human breath to detect illness					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		1.58*	0.26	0.56	0.29
Familiarity	110	0.03	0.06	0.08	0.06
Cognitive Trust	110	0.63*	0.08	0.38*	0.08
Benevolence	110			0.46*	0.08
R^2		0.44		0.58*	
R^2 Change				0.14*	

HC7 Helping farmers remove weeds and collect the harvest					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		1.98*	0.27	1.39*	0.30
Familiarity	111	-0.13*	0.05	-0.14*	0.05
Cognitive Trust	111	0.67*	0.08	0.50*	0.08
Benevolence	111			0.32*	0.08
R^2		0.43		0.50*	
R^2 Change				0.07*	

HC8 Discovering new chemical molecules for pharmaceutical or industrial applications					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.94*	0.30	0.63*	0.31
Familiarity	112	0.09	0.05	0.05	0.05
Cognitive Trust	112	0.71*	0.08	0.58*	0.09
Benevolence	112			0.24*	0.09
R^2		0.49		0.52*	
R^2 Change				0.04*	

HC9 Checking large volumes of documents for relevant legal evidence					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.94*	0.27	0.48*	0.27
Familiarity	112	0.17*	0.05	0.11*	0.05
Cognitive Trust	112	0.66*	0.07	0.54*	0.07
Benevolence	112			0.29*	0.06
R^2		0.56		0.63*	
R^2 Change				0.07*	

HC10 Reducing fraud related to exams or assessments					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.65	0.35	0.21	0.35
Familiarity	112	0.12	0.06	0.07	0.06
Cognitive Trust	113	0.71	0.09	0.58*	0.09
Benevolence	113			0.31*	0.08
R^2		0.43		0.50*	
R^2 Change				0.07*	

Table 4. Continued

LC1 Predicting relationship breakdowns via listening smart home devices					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.18	0.25	-0.12	0.21
Familiarity	111	0.61	0.08	0.40*	0.07
Cognitive Trust	111	0.27	0.09	0.08	0.08
Benevolence	111			0.52*	0.08
R^2		0.59		0.72*	
R^2 Change				0.13*	
LC2 Providing psychological counseling					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.37	0.20	-0.02	0.21
Familiarity	111	0.19*	0.08	0.12	0.08
Cognitive Trust	111	0.63*	0.09	0.42*	0.10
Benevolence	111			0.38*	0.08
R^2		0.58		0.64*	
R^2 Change				0.07*	
LC3 Being a primary care physician (Acting as a doctor in a general practice)					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.25	0.20	0.04	0.20
Familiarity	110	0.41*	0.08	0.36*	0.08
Cognitive Trust	110	0.50*	0.08	0.36*	0.09
Benevolence	110			0.24*	0.07
R^2		0.64		0.67*	
R^2 Change				0.03*	
LC4 Selecting staff for employment					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		-0.22	0.26	-0.27	0.25
Familiarity	111	0.19*	0.08	0.16*	0.08
Cognitive Trust	111	0.80*	0.09	0.67*	0.11
Benevolence	111			0.18*	0.09
R^2		0.61		0.62*	
R^2 Change				0.01*	
LC5 Playing a team football match					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		1.00*	0.21	0.97*	0.21
Familiarity	110	0.19	0.10	-0.04	0.16
Cognitive Trust	110	0.52*	0.10	0.40*	0.12
Benevolence	110			0.34	0.18
R^2		0.46		0.48*	
R^2 Change				0.02	

Table 4. Continued

LC6 Being a news anchor					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.75*	0.26	0.53*	0.26
Familiarity	111	0.22*	0.07	0.12	0.07
Cognitive Trust	111	0.56*	0.08	0.45*	0.08
Benevolence	111			0.30*	0.09
R^2		0.47		0.52*	
R^2 Change				0.05*	
LC7 Being an actor in a film					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		1.11*	0.21	1.02*	0.22
Familiarity	111	0.07	0.09	-0.04	0.11
Cognitive Trust	111	0.65*	0.10	0.61*	0.10
Benevolence	111			0.18	0.12
R^2		0.51		0.52*	
R^2 Change				0.01	
LC8 Being a bank branch employee					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.86*	0.35	0.42	0.33
Familiarity	111	0.15*	0.08	0.01	0.08
Cognitive Trust	111	0.62*	0.10	0.49*	0.10
Benevolence	111			0.40*	0.08
R^2		0.36		0.47*	
R^2 Change				0.11*	
LC9 Driving a car					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.13	0.35	-0.20	0.32
Familiarity	111	0.12	0.09	0.00	0.08
Cognitive Trust	111	0.77*	0.09	0.50*	0.10
Benevolence	111			0.49*	0.09
R^2		0.48		0.60*	
R^2 Change				0.12*	
LC10 Deciding how to prioritize aid during a humanitarian crisis					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant		0.30	0.21	0.05	0.21
Familiarity	114	0.13*	0.05	0.10	0.05
Cognitive Trust	113	0.76*	0.07	0.58*	0.08
Benevolence	114			0.27*	0.07
R^2		0.66		0.70*	
R^2 Change				0.04*	

*$p < .05$
Note. Unstandardized *B*s are reported. HC = High Comfort
HC= High Comfort; LC = Low Comfort

Table 5.

AI Rankings for 1b Selection

	Artificial Intelligence Application	R^2 Δ	Ben. Std. *B*	Affective Trust *M*	total score	Rank
HC5	**Forecasting storm damage in forestry plantations**	**1**	**1**	**5**	**7**	**1**
HC6	**Using smells in human breath to detect illness**	**2**	**4**	**8**	**14**	**2**
HC3	**Searching for life on other planets**	**6**	**8**	**1**	**15**	**3**
LC9	**Driving a car**	**4**	**3**	**11**	**18**	**4**
HC7	Helping farmers remove weeds and collect the harvest	8	9	2	19	5
HC2	Teaching people sign language	7	7	6	20	6
LC8	Being a bank branch employee	5	5	10	20	7
LC1	Predicting relationship breakdowns via listening smart home devices	3	2	17	22	8
HC9	Checking large volumes of documents for relevant legal evidence	9	10	4	23	9
HC10	Reducing fraud related to exams or assessments	10	11	9	30	10
LC2	Providing psychological counseling	11	6	15	32	11
HC4	Working in car manufacturing plants	16	16	3	35	12
HC8	Discovering new chemical molecules for pharmaceutical or industrial applications	14	15	7	36	13
LC6	Being a news anchor	12	12	13	37	14
LC10	Deciding how to prioritize aid during a humanitarian crisis	13	13	12	38	15
LC3	Being a primary care physician (Acting as a doctor in a general practice)	15	15	16	46	16
LC4	Selecting staff for employment	17	17	14	48	17

Table 6.

Study 1b: Means, Standard Deviations, Intercorrelations, and Cronbach Alphas

AI		*M*	*SD*	1	2	3	4
Forecasting storm damage in forestry plantations	1. Cognitive Trust	3.98	0.67	**(0.84)**			
	2. Benevolence	4.13	0.59	0.61*	**(0.87)**		
	3. Familiarity	3.24	1.27	0.46*	0.21*	**(0.92)**	
	4. Affective Trust	4.06	0.64	0.73*	0.69*	0.23*	**(0.82)**
		M	*SD*	1	2	3	**4**
Using smells in human breath to detect illness	1. Cognitive Trust	3.66	0.88	**(0.91)**			
	2. Benevolence	4.18	0.62	0.17*	**(0.89)**		
	3. Familiarity	3.01	1.38	0.67*	-0.09	**(0.94)**	
	4. Affective Trust	3.93	0.71	0.56*	0.50*	0.35*	**(0.83)**
		M	*SD*	1	2	3	**4**
Searching for life on other planets	1. Cognitive Trust	3.97	0.77	**(0.85)**			
	2. Benevolence	3.83	0.79	0.67*	**(0.90)**		
	3. Familiarity	3.37	1.26	0.55*	0.57*	**(0.91)**	
	4. Affective Trust	4.10	0.68	0.74*	0.63*	0.34*	**(0.84)**
		M	*SD*	1	2	3	**4**
Driving a Car	1. Cognitive Trust	3.75	0.91	**(0.91)**			
	2. Benevolence	3.82	0.85	0.79*	**(0.92)**		
	3. Familiarity	3.87	0.88	0.32*	0.39*	**(0.75)**	
	4. Affective Trust	3.63	1.00	0.87*	0.84*	0.26*	**(0.92)**

*$p < 0.05$

Note. N=161; Cronbach's alphas are shown on the diagonal.

Table 7.

Study 1b: Confirmatory Factor Analysis

Forecasting storm damage in forestry plantations

Model	X^2	*df*	ΔX^2	Δdf	RMSEA	CFI	SRMR
4-Factor Model	319.82*	183			0.07	0.92	0.05
3-Factor Model[1]	355.79*	186	35.97*	3	0.08	0.90	0.06
1-Factor Model[2]	762.99*	189	443.17*	6	0.14	0.67	0.10

Using smells in human breath to detect illness

Model	X^2	*df*	ΔX^2	Δdf	RMSEA	CFI	SRMR
4-Factor Model	277.47*	183			0.06	0.96	0.06
3-Factor Model[1]	480.66*	186	203.19*	3	0.10	0.86	0.11
1-Factor Model[2]	1222.42*	189	944.95*	6	0.18	0.51	0.20

Searching for life on other planets

Model	X^2	*df*	ΔX^2	Δdf	RMSEA	CFI	SRMR
4-Factor Model	372.96*	183			0.08	0.91	0.06
3-Factor Model[1]	413.35*	186	40.39*	3	0.09	0.89	0.07
1-Factor Model[2]	798.05*	189	425.09*	6	0.14	0.71	0.09

Driving a car

Model	X^2	*df*	ΔX^2	Δdf	RMSEA	CFI	SRMR
4-Factor Model	344.21*	183			0.07	0.94	0.04
3-Factor Model[1]	368.50*	186	24.29*	3	0.08	0.93	0.04
1-Factor Model[2]	545.75*	189	201.54*	6	0.11	0.86	0.07

Note. [1]Combined cognitive trust and affective trust, [2]Combined all items.

**p* < .001 level.

Table 8.

Study 1b: Hierarchical Regression Analysis for DV Affective Trust

Forecasting storm damage in forestry plantations					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant	161	1.28*	0.21	0.47*	0.23
Familiarity	161	-0.08*	0.03	-0.05	0.03
Cognitive Trust	161	0.75*	0.06	0.52*	0.06
Benevolence	161			0.41*	0.07
R^2		0.55*		0.63*	
R^2 Change				0.09*	

Using smells in human breath to detect illness					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant	161	2.27*	0.20	0.43	0.30
Familiarity	161	-0.02	0.05	0.06	0.04
Cognitive Trust	161	0.47*	0.07	0.32*	0.07
Benevolence	161			0.51*	0.07
R^2		0.31		0.49*	
R^2 Change				0.18*	

Searching for life on other planets					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant	161	1.50*	0.19	1.19*	0.19
Familiarity	161	-0.05	0.03	-0.10*	0.03
Cognitive Trust	161	0.70*	0.06	0.56*	0.06
Benevolence	161			0.27*	0.06
R^2		0.55		0.60*	
R^2 Change				0.05*	

Driving a car					
		Step 1		Step 2	
	N	*B*	*SE*	*B*	*SE*
Constant	161	0.12	0.21	-0.22	0.18
Familiarity	161	-0.02	0.05	-0.10*	0.04
Cognitive Trust	161	0.96*	0.05	0.60*	0.06
Benevolence	161			0.52*	0.07
R^2		0.76		0.83*	
R^2 Change				0.07*	

Note. Unstandardized *B*s are reported.
*$p < 0.05$

Table 9.

Study 2: Manipulation of Artificial Intelligence and Augmented Intelligence Descriptions

	Artificial Intelligence	Augmented Intelligence
#1 AI	Artificial Intelligence that predicts relationship breakdowns via listening smart home devices.	Augmented Intelligence that collaborates with couples to predict relationship breakdowns via listening smart home devices.
#2 AI	Artificial Intelligence that offers patients primary care diagnoses and treatments.	Augmented Intelligence that collaborates with doctors to offer patients primary care diagnoses and treatments.
#3 AI	Artificial Intelligence that provides psychological counseling.	Augmented Intelligence that collaborates with counselors to provide psychological counseling.
#4 AI	Artificial Intelligence that selects staff for employment.	Augmented Intelligence that collaborates with recruiters to select staff for employment.
#5 AI	Artificial Intelligence that acts as a news anchor.	Augmented Intelligence that collaborates with journalists to act as a news anchor.
#6 AI	Artificial Intelligence that decides how to prioritize aid during a humanitarian crisis.	Augmented Intelligence that collaborates with humanitarian workers to decide how to prioritize aid during a humanitarian crisis.
#7 AI	Artificial Intelligence that drives a car.	Augmented Intelligence that collaborates with drivers to drive a car.
#8 AI	Artificial Intelligence that acts as a bank branch employee.	Augmented Intelligence that collaborates with bank employees to act as a bank branch employee.
#9 AI	AI that reduces fraud related to exams.	Augmented Intelligence that collaborates with exam graders to reduce fraud related to exams.
#10 AI	AI that uses smells in human breath to detect illness.	Augmented Intelligence that collaborates with doctors to use smells in human breath to detect illness.

Table 10.

Study 2: Means, Standard Deviations, and Intercorrelations

#1 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	201	3.60	1.02	0.21*	1			
3. Benevolence	203	3.45	1.14	0.03	0.55*	1		
4. Familiarity	202	3.29	1.38	0.06	0.66*	0.62*	1	
5. Affective Trust	203	3.60	1.07	0.07	0.31*	0.33*	0.40*	1
#2 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.83	0.98	-0.03	1			
3. Benevolence	203	3.92	0.89	-0.02	0.40*	1		
4. Familiarity	203	3.52	1.06	0.08	0.41*	0.25*	1	
5. Affective Trust	203	3.60	1.07	0.07	0.51*	0.32*	0.46*	1
#3 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.55	1.06	0.01	1			
3. Benevolence	202	3.71	0.97	0.02	0.42*	1		
4. Familiarity	203	3.35	1.25	0.02	0.50*	0.29*	1	
5. Affective Trust	203	3.38	1.20	0.03	0.63*	0.43*	0.47*	1
#4 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.79	0.96	0.02	1			
3. Benevolence	203	3.55	1.00	-0.02	0.44*	1		
4. Familiarity	202	3.50	1.03	0.06	0.37*	0.44*	1	
5. Affective Trust	202	3.54	1.12	0.04	0.60*	0.37*	0.35*	1.00
#5 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.81	1.03	-0.09	1			
3. Benevolence	203	3.54	1.10	-0.12	0.50*	1		
4. Familiarity	203	3.38	1.21	0.02	0.47*	0.55*	1	
5. Affective Trust	203	3.60	1.06	-0.07	0.57*	0.38*	0.43*	1
#6 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.77	0.96	0.00	1			
3. Benevolence	203	3.85	0.91	0.05	0.30*	1		
4. Familiarity	203	3.35	1.25	-0.07	0.41*	0.30*	1	
5. Affective Trust	203	3.50	1.04	0.02	0.61*	0.41**	0.33*	1

Table 10. Continued

#7 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.85	0.86	0.03	1			
3. Benevolence	202	3.77	0.92	0.06	0.34*	1		
4. Familiarity	202	3.80	0.91	0.13	0.17*	0.40*	1	
5. Affective Trust	203	3.58	0.98	0.01	0.55*	0.40*	0.22*	1
#8 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.92	0.86	-0.05	1			
3. Benevolence	203	3.71	0.96	0.04	0.37*	1		
4. Familiarity	203	3.51	1.11	-0.04	0.44*	0.55*	1	
5. Affective Trust	203	3.73	0.92	-0.03	0.54*	0.28*	0.32*	1
#9 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.93	0.86	0.03	1			
3. Benevolence	203	3.73	0.93	0.10	0.26*	1		
4. Familiarity	202	3.48	1.16	-0.03	0.12	0.29*	1	
5. Affective Trust	203	3.84	0.95	0.00	0.40*	0.31*	0.05	1
#10 AI	*N*	*M*	*SD*	1	2	3	4	5
1. Group	203	0.51	0.50	1				
2. Cognitive Trust	203	3.75	0.89	0.07	1			
3. Benevolence	202	3.85	1.00	-0.06	0.36*	1		
4. Familiarity	203	3.30	1.26	0.10	0.21*	0.08	1	
5. Affective Trust	202	3.80	1.08	-0.01	0.36*	0.42*	0.14*	1

*$p < 0.05$

Table 11.

Study 2: Mediation Results Table

#1 Predicts Relationship Breakdowns	Benevolence		Affective Trust	
	B	*SE*	*B*	*SE*
Constant	1.20*	.23	.15	.26
Group	-.12	.13	-.08	.14
Cognitive Trust	.30*	.06	.31*	.09
Familiarity	.37*	.06	.30*	.07
Benevolence			.32*	.08
R^2	.42*		.51*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	-.04	.04	-.13	.04
#2 Provides Diagnosis & Treatment	**Benevolence**		**Affective Trust**	
	B	*SE*	*B*	*SE*
Constant	2.39*	.26	.52	.33
Group	-.03	.12	.12	.12
Cognitive Trust	.32*	.06	.34*	.07
Familiarity	.09	.06	.29*	.06
Benevolence			.14	.08
R^2	.17*		.35*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	-.00	.02	-.12	.37
#3 Provides Psychological Counseling	**Benevolence**		**Affective Trust**	
	B	*SE*	*B*	*SE*
Constant	2.21*	.24	.06	.29
Group	.04	.12	.05	.13
Cognitive Trust	.34*	.07	.53*	.07
Familiarity	.08	.06	.17*	.06
Benevolence			.23*	.07
R^2	.19*		.46*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	.01	.03	-.05	.07
#4 Selects Staff for Employment	**Benevolence**		**Affective Trust**	
	B	*SE*	*B*	*SE*
Constant	1.24*	.28	.44	.30
Group	-.09	.12	.04	.13
Cognitive Trust	.33*	.07	.61*	.07
Familiarity	.32*	.06	.13	.07
Benevolence			.09	.07
R^2	.29*		.39*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	-.01	.02	-.05	.02

Table 11. Continued

#5 Acts as a News Anchor	Benevolence		Affective Trust	
	B	*SE*	*B*	*SE*
Constant	1.17*	.26	1.09*	.27
Group	-.21	.12	-.05	.12
Cognitive Trust	.31*	.07	.48*	.07
Familiarity	.38*	.06	.16*	.06
Benevolence			.05	.07
R^2	.39*		.36*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	-.01	.02	-.07	.19
#6 Prioritizes Aid During a Crisis	**Benevolence**		**Affective Trust**	
	B	*SE*	*B*	*SE*
Constant	2.50*	.26	.18	.29
Group	.11	.12	.03	.11
Cognitive Trust	.20*	.07	.57*	.07
Familiarity	.16*	.05	.03	.05
Benevolence			.27*	.07
R^2	.13*		.43*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	.03	.03	-.19	.25
#7 Drives a Car	**Benevolence**		**Affective Trust**	
	B	*SE*	*B*	*SE*
Constant	1.27*	.33	.44	.33
Group	.01	.12	-.07	.11
Cognitive Trust	.30*	.07	.53*	.07
Familiarity	.36*	.06	.07	.07
Benevolence			.23*	.07
R^2	.23*		.36*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	.00	.03	-.05	.05
#8 Acts as a Bank Branch Employee	**Benevolence**		**Affective Trust**	
	B	*SE*	*B*	*SE*
Constant	1.50*	.28	1.29*	.29
Group	.12	.11	-.02	.11
Cognitive Trust	.18*	.07	.52*	.07
Familiarity	.41*	.05	.06	.06
Benevolence			.06	.07
R^2	.32*		.30*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	.01	.01	-.01	.04

Table 11. Continued

#9 Reduces Fraud Related to Exams	Benevolence		Affective Trust	
	B	*SE*	*B*	*SE*
Constant	2.00*	.33	1.59*	.36
Group	.17	.12	-.07	.12
Cognitive Trust	.23*	.07	.39*	.07
Familiarity	.21*	.05	-.06	.06
Benevolence			.26*	.07
R^2	.14*		.21*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	.04	.04	-.02	.13

#10 Uses Smells to Detect Illness	Benevolence		Affective Trust	
	B	*SE*	*B*	*SE*
Constant	2.38*	.31	1.21*	.36
Group	-.18	.13	-.02	.36
Cognitive Trust	.40*	.08*	.28*	.08
Familiarity	.01	.05	.05	.06
Benevolence			.36*	.07
R^2	.13*		.23*	
Indirect Effect	Effect	Boot *SE*	LCI	UCI
Group → Benevolence → Affective Trust	-.07	.05	-.18	.03

**p* < 0.05. *B*s are unstandardized.

Table 12.

Study 3: Benevolence Messaging Manipulation

No manipulation	AI benevolence manipulation
AI that acts as a Primary Care Doctor This AI reviews patient data including history and current symptoms to diagnosis and recommend treatment.	AI that acts as a Primary Care Doctor This AI reviews patient data including history and current symptoms to diagnosis and recommend treatment. This AI is intended to promote human welfare by reducing medical errors from incomplete assessments which currently cause about 35% of medical errors.
AI that selects staff for employment This AI scans information to find staff with skills desired for job.	AI that selects staff for employment This AI scans information to find staff with skills desired for job. This AI is intended to promote human welfare by fairly comparing candidates' skills to reduce discriminatory hiring. Currently, biases of recruiters regularly lead to unfair hires which also causes productivity losses.
AI that acts as a bank branch employee This AI conducts routine banking transactions.	AI that acts as a bank branch employee This AI conducts routine banking transactions. This AI is intended to promote human welfare by conducting routine banking transactions more securely with voice and face recognition. Currently fraudulent banking is up 10% compared to before the pandemic.
AI that drives a car This AI continuously senses the environment around the car and controls the car responses.	AI that drives a car This AI continuously senses the environment around the car and controls the car responses. This AI is intended to promote human welfare by reducing human errors which currently cause 94% of traffic deaths.
AI that prioritizes aid during a humanitarian crisis This AI determines recipients of aid and aid routes.	AI that prioritizes aid during a humanitarian crisis This AI determines recipients of aid and aid routes. This AI is intended to promote human welfare by predicting the food needs of individual communities supporting refugee populations. In the last 10 years, 2 billion people have been displaced from their homes due to weather or political turmoil.

Table 13.

Study 3: Means, Standard Deviations, and Intercorrelations

#1 Acts as Primary Care Doctor	*N*	*M*	*SD*	1	2	3	4	5	6	7
1. Group	199	0.51	0.50	1						
2. Likelihood of Worker Replacement	199	3.69	1.03	0.14	1					
3. Likelihood of Death from Error	199	3.66	1.06	-0.02	0.16*	1				
4. Familiarity	199	3.14	1.35	0.11	0.31*	-0.01	1			
5. Cognitive Trust	199	3.49	1.04	0.05	0.24*	0.00	0.59*	1		
6. Benevolence	199	3.54	1.00	0.03	0.19*	0.06	0.48*	.50*	1	
7. Affective Trust	199	3.25	1.14	0.10	0.19*	-0.09	0.57*	.66*	.38*	1
#2 Selects Staff for Employment	*N*	*M*	*SD*	1	2	3	4	5	6	7
1. Group	199	0.51	0.50	1						
2. Likelihood of Worker Replacement	199	3.79	0.95	-0.04	1					
3. Likelihood of Death from Error	199	2.82	1.45	0.03	0.14*	1				
4. Familiarity	198	3.36	1.12	-0.05	0.28*	0.46*	1			
5. Cognitive Trust	199	3.70	0.99	-0.16*	0.19*	0.32*	0.47*	1		
6. Benevolence	199	3.47	1.04	-0.07	0.18*	0.24**	0.43*	.60*	1	
7. Affective Trust	199	3.57	1.05	-0.1	0.13	0.28*	0.45*	.59*	.54*	1
#3 Acts as Bank Branch Employee	*N*	*M*	*SD*	1	2	3	4	5	6	7
1. Group	199	0.51	0.50	1						
2. Likelihood of Worker Replacement	199	3.87	0.95	-0.09	1					
3. Likelihood of Death from Error	197	2.84	1.46	0.01	-0.04	1				
4. Familiarity	198	3.28	1.18	-0.13	0.16*	0.48*	1			
5. Cognitive Trust	199	3.84	0.85	-0.11	0.35*	0.05	0.27*	1		
6. Benevolence	197	3.49	0.88	0.07	0.11	0.31*	0.44*	0.33*	1	
7. Affective Trust	198	3.75	0.96	-0.09	0.25*	0.10	0.29*	0.49*	0.26*	1

Table 13. Continued

#4 Drives a Car	*N*	*M*	*SD*	1	2	3	4	5	6	7
1. Group	199	0.51	0.50	1						
2. Likelihood of Worker Replacement	199	3.65	1.15	0	1					
3. Likelihood of Death from Error	199	3.77	1.00	0.07	0.10	1				
4. Familiarity	199	3.43	1.07	0.11	0.35*	.141*	1			
5. Cognitive Trust	199	3.71	0.86	-0.09	0.28*	0.04	0.37*	1		
6. Benevolence	199	3.59	0.93	0.01	0.34*	-0.02	0.45*	0.39*	1	
7. Affective Trust	198	3.44	1.08	-0.04	0.26*	-0.16*	0.36*	0.52*	0.40*	1
#5 Prioritizes Aid During Crisis	*N*	*M*	*SD*	1	2	3	4	5	6	7
1. Group	199	0.51	0.50	1						
2. Likelihood of Worker Replacement	199	3.58	1.01	0.00	1					
3. Likelihood of Death from Error	199	3.48	1.09	-0.02	0.29*	1				
4. Familiarity	199	3.04	1.39	0.10	0.36*	0.18*	1			
5. Cognitive Trust	199	3.65	0.90	0.05	0.33*	0.14	0.44*	1		
6. Benevolence	198	3.69	0.94	0.14*	0.21*	0.02	0.22*	0.37*	1	
7. Affective Trust	199	3.58	1.06	0.16*	0.21*	0.03	0.46*	0.57*	0.37*	1

*$p < 0.05$

Table 14.

Study 3: Results with Framing as Moderator and Affective Trust as DV

#1 Primary Care Doctor	*B*	*SE*
Constant	0.56*	0.22
Likelihood of Worker Replacement	0.00	0.08
Likelihood of Death from Error	-0.04	0.09
Group	0.10	0.12
Familiarity	0.24*	0.06
Cognitive Trust	0.54*	0.07
Likelihood of Worker Replacement × Group	-0.04	0.12
Likelihood of Death from Error × Group	-0.09	0.11
R^2	0.50*	
#2 Selects Staff	*B*	*SE*
Constant	0.83*	0.30
Likelihood of Worker Replacement	0.07	0.09
Likelihood of Death from Error	-0.02	0.06
Group	0.00	0.12
Familiarity	0.23*	0.07
Cognitive Trust	0.53*	0.07
Likelihood of Worker Replacement × Group	-0.22	0.13
Likelihood of Death from Error × Group	0.07	0.08
R^2	0.41*	
#3 Bank Branch Employee	*B*	*SE*
Constant	1.49*	0.34
Likelihood of Worker Replacement	0.09	0.10
Likelihood of Death from Error	-0.01	0.06
Group	-0.03	0.12
Familiarity	0.13*	0.06
Cognitive Trust	0.48*	0.08
Likelihood of Worker Replacement × Group	-0.01	0.13
Likelihood of Death from Error × Group	0.03	0.08
R^2	0.28*	
#4 Drives a Car	*B*	*SE*
Constant	0.76*	0.33
Likelihood of Worker Replacement	0.10	0.08
Likelihood of Death from Error	-0.19	0.10
Group	-0.02	0.13
Familiarity	0.22*	0.07
Cognitive Trust	0.53*	0.08
Likelihood of Worker Replacement × Group	-0.04	0.11
Likelihood of Death from Error × Group	-0.05	0.13
R^2	0.35*	

Table 14. Continued

#5 Prioritizes Aid	*B*	*SE*
Constant	0.84*	0.27
Likelihood of Worker Replacement	0.04	0.09
Likelihood of Death from Error	-0.10	0.08
Group	0.36	0.52
Familiarity	0.21*	0.05
Cognitive Trust	0.55*	0.07
Likelihood of Worker Replacement × Group	-0.10	0.13
Likelihood of Death from Error × Group	0.07	0.12
R^2	0.41*	

*$p < 0.05$. *B*s are unstandardized.

Table 15.

Study 3: Modification Results with Benevolence as Modifier and Affective Trust as DV

#1 Primary Care Doctor		*B*	*SE*
Constant		0.62*	0.25
Likelihood of Worker Replacement		-0.01	0.06
Likelihood of Death from Error		-0.09	0.06
Benevolence		0.00	0.07
Familiarity		0.24*	0.06
Cognitive Trust		0.54*	0.07
Likelihood of Worker Replacement × Benevolence		0.01	0.05
Likelihood of Death from Error × Benevolence		0.00	0.05
	R2	0.49*	
#2 Selects Staff		***B***	***SE***
Constant		1.68*	0.32
Likelihood of Worker Replacement		-0.02	0.06
Likelihood of Death from Error		0.02	0.05
Benevolence		0.24*	0.07
Familiarity		0.17*	0.06
Cognitive Trust		0.36*	0.08
Likelihood of Worker Replacement × Benevolence		0.01	0.05
Likelihood of Death from Error × Benevolence		-0.10*	0.04
	R2	0.46*	
#3 Bank Branch Employee		***B***	***SE***
Constant		1.58*	0.34
Likelihood of Worker Replacement		0.16*	0.07
Likelihood of Death from Error		0.00	0.05
Benevolence		0.01	0.08
Familiarity		0.14*	0.06
Cognitive Trust		0.46*	0.08
Likelihood of Worker Replacement × Benevolence		0.06	0.06
Likelihood of Death from Error × Benevolence		-0.14*	0.05
	R2	0.33*	
#4 Drives a Car		***B***	***SE***
Constant		1.20*	0.35
Likelihood of Worker Replacement		0.06	0.06
Likelihood of Death from Error		-0.18*	0.06
Benevolence		0.17*	0.08
Familiarity		0.16*	0.07
Cognitive Trust		0.46*	0.08
Likelihood of Worker Replacement × Benevolence		-0.05	0.05
Likelihood of Death from Error × Benevolence		0.12	0.07
	R2	0.38*	

Table 15. Continued

#5 Prioritizes Aid	*B*	*SE*
Constant	1.14*	0.28
Likelihood of Worker Replacement	-0.04	0.07
Likelihood of Death from Error	-0.07	0.06
Benevolence	0.21*	0.07
Familiarity	0.20*	0.05
Cognitive Trust	0.50*	0.08
Likelihood of Worker Replacement × Benevolence	0.08	0.06
Likelihood of Death from Error × Benevolence	-0.01	0.06
R2	0.42*	

*p < 0.05. *B*s are unstandardized.

APPENDIX B

STUDY 1B CONSTRUCT MEASUREMENT SCALES

Cognitive Trust

Adapted from Mayer and Davis's Capability Scale (1999)

CT1	I think AI is very capable of doing this task.
CT2	I think AI is known to be successful at doing this task.
CT4	I am very confident about AI's skills for doing this task.
CT5	AI has specialized capability that can increase performance of doing this task.
CT6	I think AI is well-qualified for doing this task.

Benevolence

Adapted from Mayer and Davis (1999)

BN1	AI doing this task is in humans' welfare.
BN2	People's needs are served by AI that does this task.
BN3	AI that does this task will keep people safe.
BN4	AI that does this task is important to people.
BN5	AI that does this task will help people.

Benevolence New Items

BN6	AI doing this task is in humans' best interest.
BN7	AI doing this task is good for humans.

Familiarity

Adapted from Gefen (2000)

FM1	I am familiar with AI doing (task).
FM2	I am familiar with the computer processes likely used to develop AI for this task.
FM3	I am familiar with applying AI to do (task).

Affective Trust

Adapted from Mayer and Davis (1999)

AT1	I feel comfortable with AI doing this task.
AT2	I feel comfortable giving AI complete responsibility for doing this task.
AT4	I feel comfortable allowing AI to implement its recommended action even if I could not monitor it.

Affective Trust New Items

AT5	I am optimistic about AI being used for doing this task.
AT6	AI that can perform this task is exciting.
AT7	I feel positive about AI doing this task.

VITA

Marjorie A. Erdmann

Candidate for the Degree of

Doctor of Philosophy

Thesis: UNDERSTANDING AFFECTIVE TRUST IN AI: THE EFFECTS OF PERCEIVED BENEVOLENCE

Major Field: Business Administration

Education:

Completed the requirements for the Doctor of Philosophy in Business Administration at Oklahoma State University, Stillwater, Oklahoma in December, 2022.

Completed the requirements for the Master of Science in Family Relations and Child Development, Marriage and Family Therapy at Oklahoma State University, Stillwater, OK in December 1994.

Completed the requirements for the Bachelor of Science in Sociology at DePauw University, Greencastle, IN, in May 1991.

Experience:

2014 – Present, Director Center for Health Systems Innovation, Oklahoma State University, Stillwater, OK
2011 – 2014, Director of Business Development, UHS, Inc., Tulsa, OK
1999 – 2002, Pharma Sales, Novartis Pharmaceuticals, Tulsa, OK
1998 – 1999, Regional Representative, Premier Health Services, Tulsa, OK
1995 – 1998, Coordinator of School Services CMC, Tulsa

Professional Memberships:

American Academy of Management
Southern Management Association