TRUST AND DISTRUST IN ARTIFICIAL INTELLIGENCE (AI) AGENTS: A CONSTRUAL-LEVEL PERSPECTIVE

Liqiong Deng, University of West Georgia

ABSTRACT

As Artificial Intelligence (AI) technologies are becoming ubiquitous in the modern world, AI agents have been increasingly adopted to serve various roles in our daily lives, such as personal assistant, salesperson, customer service agent, and virtual counselor. Thus, interacting with AI agents has become an everyday activity, which has received much research attention. Addressing the need to understand the interaction between humans and AI, this paper develops a research model of how user experience with AI agents influences users' trust and distrust in AI agents. More specifically, it categorizes the attributes of user experience with AI agents as process-related vs outcome-related. Drawing on the two-factor theory, construal level theory (CLT), and IS success and AI trust research, the research model proposes the differential effects of process-related and outcome-related attributes of AI user experience on users' trusting and distrusting beliefs in AI agents that are moderated by users' construal levels as well as the subsequent effects of trusting and distrusting beliefs on continued intention to use AI agents. In addition, the research model suggests that a construal fit between users' perception of AI agents and their AI usage context will increase their trusting belief in AI agents. By providing an understanding of the role of construal fit in promoting trust and the psychological mechanism by which various attributes of AI user experience differentially influence users' trust and distrust in AI agents, this paper will offer guidelines on how to appropriately design and implement AI agents to enhance trust and minimize distrust.

Keywords: artificial intelligence, construal level, trust, distrust

INTRODUCTION

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think and act like humans (Bansal, Pruthi & Singh, 2022). AI Agents are intelligent nonhuman agents (Kim & Duhachek, 2020) that have been increasingly adopted to serve various roles in our daily lives, such as personal assistants, salespersons, customer service agents, virtual counselors, and etc. AI agents encompass a fusion of mathematical algorithms and systems designed to interact with humans and conduct tasks that were previously only performed by humans (Jackson, 2019; Kim & Duhachek, 2020). AI agents are capable of acting, reacting and cooperating much like humans do (Russell & Norvig, 2009). AI has been developed from a narrow concept of intelligence focused on task performance to a broader concept known as super AI, which is expected to integrate social skills and more complex problem-solving and creation (Kaplan & Haenlein, 2019). One key distinguishing factor between automated IT systems and AI agents lies in the AI agents' ability to learn and independently recommend solutions to humans (Szolovits, 2019).

Despite the advances in AI technologies and their applications across a variety of platforms and applications, research has shown many people are still averse to interactions with AI agents

(Dietvorst, Simmons & Massey, 2015). This reluctance could be due to the uncanny valley effect (Mori, MacDorman & Kageki, 2012), job displacement fears (Frey & Osborne, 2017), privacy and security concerns (Bélanger & Crossler, 2011), challenges of establishing emotional connections with AI (Picard, 1997), and lack of trust in AI decision-making (Lee & See, 2004). This research will focus on investigating the factors influencing people's trust and distrust in AI agents that have important implications for improving users' interactions with AI agents and increasing intentions to use AI agents. Prior research suggests that user experience is closely related to users' subjective evaluations of the interaction with a system (Hassenzahl & Tractinsky, 2006; Knijnenburg, Willemsen, Ganter, Soncu & Newell, 2012). User experience plays a crucial role in shaping users' trust and distrust in AI agents and their continued intention to use such agents. For instance, a positive user experience can foster trust and encourage continued usage, while a negative one can engender distrust and discourage further usage. So, this research aims to develop a research model of how user experience with AI agents influences users' trust and distrust in AI agents as well as their continued intention to use AI agents. By understanding how users evaluate various attributes of their experience with AI agents to form trusting and distrusting beliefs as well as continuous intention in AI agents, this research can guide the development of AI agents that maximize user trust, minimize distrust, and promote sustained usage. The remainder of this paper is structured as follows. It first reviews the existing literature on the two-factor theory, construal level theory (CLT), and IS success and AI trust research. Then, it presents the research model and its associated propositions. In the end, the expected contributions of this paper are discussed.

THEORETICAL BACKGROUND

Trust and Distrust

Lewicki, McAllister, and Bies (1998) theorize that trust and distrust are two separate constructs. This proposition is substantiated by the possibility of trust and distrust coexisting in a state of inconsistency. Consistency would entail either high trust and low distrust, or low trust and high distrust. Inconsistency, however, implies a simultaneous presence of high trust and high distrust or low trust and low distrust. Lewicki and his colleagues also argue that trust and distrust are distinct factors with unique antecedents and outcomes. Similarly, Sitkin and Roth (1993) assert that trust and distrust are distinct constructs. They define trust as confidence in an individual's ability or competence to carry out a specific task under certain circumstances. In contrast, distrust is described as the belief that a person's motives or values will lead them to behave inappropriately across situations.

Kramer (1999) also advocates the idea that trust and distrust have different antecedents and consequences. Komiak and Benbasat's (2008) work further strengthens this viewpoint by theoretically postulating and empirically verifying a dual-process perspective. They propose that the processes building trust and distrust are distinct and separate. In the context of recommendation agent (RA) usage, trust (or distrust) is the positive (or negative) expectation regarding the RA's conduct, including its interface, explanations, and recommendations. A trust-building (or distrust-building) process involves a user's favorable (or unfavorable) interpretation of interactions with an RA, leading to a positive (or negative) expectation of the RA's reliability for shopping decisions. McKnight et al. (2002) define trusting beliefs as the perceptions of positive attributes of the other party, which commonly encompass benevolence, integrity, and competence. Mayer et al. (1995)

further describe benevolence as the belief in a trustee's goodwill toward the trustor beyond self-interest. Integrity is seen as the trustee's adherence to acceptable principles, and competence is the belief in the trustee's skills and abilities to have influence within a particular domain.
Conversely, distrust is more than just low trust or lack of trust; it's the active belief that the other party will act detrimentally to one's welfare and security (Cho, 2006). Distrusting beliefs, as per Lewicki et al. (1998), are beliefs about the other party's negative attributes, including malevolence, deceit, and incompetence (Moody, Lowry & Galletta, 2010). I define malevolence, based on the work of Moody et al., (2010), as the belief that the trustee intends to cause harm, deceit as the belief in the trustee's dishonesty and potential dissemination of false information, and incompetence as the belief in the trustee's inability to complete a task.

Two-Factor Theory

The Two-Factor Theory, also known as Herzberg's Motivation-Hygiene Theory (Herzberg, Mausner & Snyderman, 1959), provides a unique lens through which one can study trusting and distrusting beliefs. Initially developed by Frederick Herzberg to analyze job satisfaction, the theory posits that factors contributing to positive attitudes (like job satisfaction) are different from those leading to negative attitudes (like job dissatisfaction) (Herzberg et al., 1959). According to the theory, motivating factors, also known as satisfiers, are elements that lead to job satisfaction and motivate employees to work harder. These factors often relate to the work itself and include aspects like achievement, recognition, work responsibility, advancement, and growth. Conversely, hygiene factors, also known as dissatisfiers, are conditions that, if absent or inadequate, cause dissatisfaction but, if met or exceeded, do not necessarily increase satisfaction. These factors typically pertain to the job environment and include company policies, supervision, working conditions, salary, and relationships with colleagues (Herzberg et al., 1959). In general, motivating factors provide individuals with a sense of achievement and enable them to experience personal growth, while hygiene factors relate to individuals' inherent drive to avoid pain from the environment (Herzberg, 1968).

Herzberg's Two-Factor Theory has traditionally been employed to study job satisfaction and dissatisfaction. However, its fundamental premise — that positive and negative attitudes are influenced by different factors — makes it applicable to other areas, including the study of trust and distrust. The Two-Factor Theory has also been used to investigate web design quality and service process quality (Johnston, 1995; Loiacono, Watson & Goodhue, 2007; Ou & Sia, 2009, 2010). Prior research shows that similar to job satisfaction and dissatisfaction, trust and distrust may also be influenced by different factors (Komiak & Benbasat, 2008; Ou & Sia, 2010). For instance, quality of communication and transparency of an AI system might enhance trust (motivating factor), while poor performance and low reliability might foster distrust (hygiene factor) (Bélanger & Crossler, 2011; Lee & See, 2004; Siau & Wang, 2018). Lewicki et al. (1998) argue that "there are elements that contribute to the growth and decline of trust, and there are elements that contribute to the growth and decline of distrust" (p. 440). This is consistent with the premise of Two-Factor Theory that certain factors will be more likely to affect positive-valent perceptions, while other factors will be more likely to affect negative-valent perceptions (Ou & Sia, 2010). Therefore, the Two-Factor Theory can serve as a valuable framework for understanding the factors that influence trusting and distrusting beliefs and provide insights for improving user experiences and trust in AI agents.

Process/Outcome Attributes of User Experience

This research focuses on the user experience with AI agents, because user experience is widely acknowledged as performing a predominant role in trust formation (Xingyuan, Li & Wei, 2010). User experience encompasses the holistic experience that a user has when interacting with a system (Hassenzahl & Tractinsky, 2006), which in this case refers to AI agents. A well-designed user experience can help foster trust and mitigate distrust (Bélanger & Crossler, 2011; Lee & See, 2004). The concept of user experience is dynamic and multi-dimensional, incorporating various evaluative aspects such as process-related attributes and outcome-related attributes (Forlizzi & Battarbee, 2004; Law, Roto, Hassenzahl, Vermeeren & Kort, 2009).

Process-related attributes refer to the elements associated with the interaction process with an AI agent aimed at accomplishing specific tasks. These attributes often include aspects like ease of use, controllability, quality of communication, sociability, and transparency. Ease of use is the simplicity/ease with which a user can navigate and interact with the system, and it is one of the most influential factors for user satisfaction and adoption (Davis, 1989). Controllability refers to the user's perceived control over the interaction with the AI agent (Novak, Hoffman & Yung, 2000). The quality of communication relates to the clarity, accuracy and relevance of the information exchanged between the user and the AI agent (Jiang, Benbasat & Wang, 2018). Sociability refers to the AI agent's ability to exhibit social behaviors that facilitate human-like interactions (Cassell, 2000). Transparency involves the AI system's ability to make its operations and decision-making processes understandable to users (Turkle, 2011).

Outcome-related attributes, on the other hand, are associated with the AI agent's performance. These include the AI's competence in completing tasks and doing so in a timely, consistent, and reliable manner (Siau & Wang, 2018). Factors such as usefulness, effectiveness, helpfulness, efficiency, and reliability come under this category (Knijnenburg, Willemsen, Ganter, Soncu & Newell, 2012; Schaefer, Chen, Szalma & Hancock, 2016). Usefulness denotes the degree to which the AI agent can enhance the user's task performance (Davis, 1989). Effectiveness refers to how accurately and successfully AI performs tasks and achieves its set goals (Haenssle et al., 2018). Helpfulness refers to the extent to which AI can assist users in achieving their goals or solving problems (VanLehn, 2011). Efficiency refers to the capability of AI to achieve its goals with minimal use of resources, such as time, energy, or computational power (Rajkomar, Dean & Kohane, 2019). And reliability is the consistency of the AI agent's performance (Parasuraman, Zeithaml & Malhotra, 2005).

Prior research indicates that process-related attributes can be classified as motivating factors, while outcome-related attributes can be regarded as hygiene factors. This classification is in line with the Two-Factor Theory (Herzberg, 1964). The process-related attributes are identified as motivating factors because a high-quality interactive process with technology leads to a more positive attitude (e.g., increased trust), while a lower quality process tends to result in a less positive attitude (e.g., decreased trust) (Beldad, de Jong & Steehouder, 2010). However, lower quality processes do not significantly influence negative attitudes, as users are generally more tolerant of a "not-so-good" interactive process than of poor outcomes (Verhagen, van Nes, Feldberg & van Dolen, 2014).

In contrast, outcome-related attributes are seen as hygiene factors because they are considered as "must-haves" and are deemed a basic and essential part of the technology (Lankton, Wilson & Mao, 2010). Users are generally unwilling to accept poor performance, such as inaccurate or incomplete results. Poor performance can lead to dissatisfaction and increased distrust in the technology (Bhattacherjee, 2001). Consequently, the outcome-related attributes have a stronger influence on distrust than on trust. Distrust increases as the AI agent's performance worsens and decreases as the performance improves (Pavlou & Gefen, 2004).

Construal Level Theory (CLT)

This research draws on the Construal-Level Theory (CLT) (Trope & Liberman, 2003, 2010) to examine how people's mental representation styles influence their evaluation of process vs outcome-related attributes of AI user experience. CLT describes how the same event or entity can be interpreted in different ways (e.g., via abstract and high-level versus concrete and low-level mental representations). These distinct interpretations influence what kind of information individuals focus on, how they process that information, and ultimately their decisions and actions regarding that event or entity. CLT is well-suited for providing insight into users' information processing style because it explains how people "make predictions, evaluations, and choices with respect to [their] construal of objects rather than the objects themselves" (Liberman & Trope, 2008, p. 1204). The level of construal that people utilize to process information shapes the kind of information they pay attention to and how they interpret it (Trope & Liberman, 2010).

The term construal refers to the way individuals perceive, comprehend, and interpret the world around them, particularly how they understand and make sense of events, actions, and objects (Trope & Liberman, 2003). It captures not only their thoughts and beliefs but also their affective, motivational, and behavioral orientations (Griffin & Ross, 1991; Mischel & Shoda, 1995). Construals may vary from highly abstract (high-level) to highly concrete (low-level) interpretations (Trope & Liberman, 2010), which profoundly alter how information is processed. High-level construals are abstract, decontextualized representations that capture the central, core features of an event or object. High-level construals are coherent and broad, and emphasize the "why" aspects - the purpose, desirability, or goal of an action or event. They are not tied to a particular context and are therefore more stable across different settings. High-level construals help individuals make sense of actions and events in a broad, overall manner without getting bogged down in the specific, incidental details. With a high-level construal, people are future oriented and focused on the desirability of distal end-states and the meaning of their actions (i.e., why actions are taken). Low-level construals, on the other hand, are concrete, detailed, and context-bound representations that include the peripheral and incidental features of an event or object. They are more focused on the "how" aspects - the means, feasibility, or process involved in an action or event. Low-level construals offer a more detailed, nuanced perspective that is attuned to the particular circumstances of a situation. So, a low-level construal "tends to contract people's mental horizons; it focuses their attention on the unique and idiosyncratic demands of present circumstances" (Wiesenfeld, Reyt, Brockner & Trope, 2017, p. 369). People using a low-level construal are oriented towards the present, vigilant in avoiding losses, and centered on the feasibility of short-term goals as well as the means for attaining them (i.e., how actions are performed) (Liberman & Trope, 1998). High- and low-level construals reflect opposing styles of information processing, and hence they are mutually exclusive.

Construal levels can be determined by individual and situational factors. Individuals may have a chronic tendency toward different levels of construal (Freitas, Salovery & Liberman, 2001; Vallacher & Wegner, 1989). As Vallacher and Wegner (1989) note, "…at one extreme is the low-level agent, someone who operates on the world primarily at the level of details. This person tends to approach an action with its mechanistic components in mind. At the other extreme is the high-level agent, someone who routinely views his or her action in terms of causal effects, social meanings, and self-descriptive implications" (p. 661). In addition, prior research has shown that psychological distance is a primary determinant of construal level (Trope & Liberman, 2003, 2010). For example, individuals tend to use detailed and concrete mental models (low-level construals) when contemplating events in the near future. Conversely, they utilize more abstract and generalized mental models (high-level construals) when considering events in the distant future (Trope & Liberman, 2000).

CLT (Trope & Liberman, 2003, 2010) also proposes a positive, reciprocal relationship between psychological distance and the level of abstraction at which a target is construed (i.e., construal level). CLT (Trope & Liberman, 2010) suggests that people represent events at varying levels of abstraction as a function of psychological distance (i.e., the removal of an event from direct experience). CLT's basic premise is that the more psychologically distant an event is, the more it will be represented at higher levels of abstraction. According to Trope and Liberman (2010), "Psychological distance refers to the perception of when an event occurs, where it occurs, to whom it occurs, and whether it occurs (p. 442)." So, there are four dimensions of psychological distance. The first dimension of temporal distance refers to the time that separates an event from the individual's present moment. The more distant the event in time, the more abstractly it is likely to be construed. The second dimension of spatial distance refers to the physical distance between an individual and an event or object. Events or objects that are far away are thought of in more abstract terms than those that are near. The third dimension of social distance involves the psychological gap between the self and others. This can pertain to interpersonal relationships (e.g., strangers versus friends) or social groups (e.g., out-groups versus in-groups). Greater social distance leads to more abstract construals. And the fourth dimension of hypotheticality refers to the likelihood of an event occurring. The less likely an event is to occur, the more psychologically distant and abstract it appears.

According to CLT, all stimuli encompass two attributes: central attributes and peripheral attributes (Liberman, Trope & Wakslak, 2007; Trope & Liberman, 2010). The central attributes of a stimulus serve to define the core characteristics or purpose of the stimulus, embodying the "what" or "why" descriptions of the stimulus, and they include features that are associated with the primary goal or actions (Higgins & Trope, 1990; Kruglanski, 1975). On the other hand, the peripheral attributes refer to the "how" description of the stimulus. They are the secondary characteristics or aspects of a stimulus, which, although not defining the essence of the stimulus, are important for its comprehensive understanding and evaluation. These attributes typically outline how a particular function or action is executed and provide additional, often contextual, details that enrich our understanding of the central entity. The "what" and "why" aspects of the stimulus correspond to the superordinate level (Liberman, Sagristano & Trope, 2002) while the "how" aspects are associated with the subordinate level of stimuli identification (Vallacher & Wegner, 1987).
As the psychological distance between the stimulus and the user – be it temporal, spatial, social, or hypothetical - changes, so does the user's evaluation of the stimulus, based on the differential

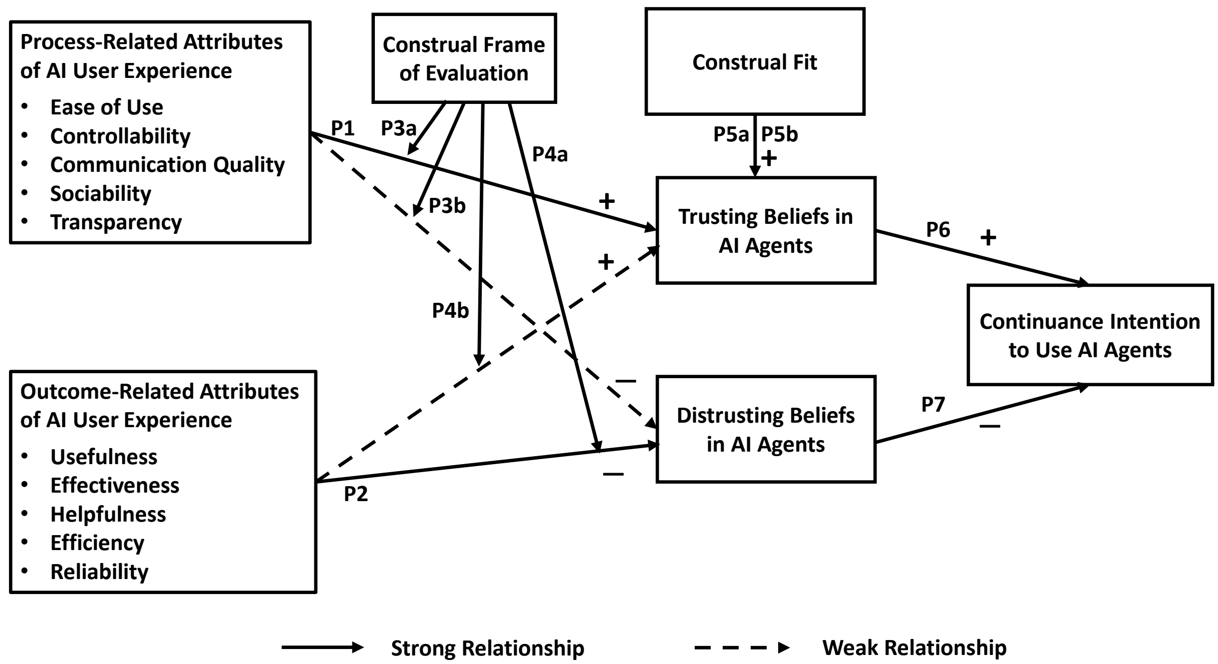
importance of the stimulus' central or peripheral features in the user's mental model (Trope, Liberman & Wakslak, 2007). An increase in distance leads to reduced importance associated with peripheral or ancillary details, shifting people's focus towards the essential or central features of the stimulus. Thus, individuals use abstract mental models to construe psychologically distant stimuli with high-level and decontextualized representations by focusing on the general core features, denoting high-level construals. Conversely, close proximity to the stimulus allows users to adopt a more detail-oriented perspective, which directs more attention to the details related to the peripheral features of the stimulus. Hence, people use concrete mental models to construe psychologically proximate stimuli with low-level and contextualized representations by concentrating on the specific detailed features, representing low-level construals. In conclusion, a high construal level is associated with the information processing of superordinate or central aspects of the stimulus, while a low construal level corresponds to the processing of subordinate or peripheral components of the stimulus (Trope & Liberman, 2000).

In terms of the user experience with AI agents, the outcome-related attributes can be regarded as central attributes, as they define the "what" and "why" of the interaction: what the AI agent can achieve and why it's useful. These are the primary characteristics that determine the value and utility of the AI agent – its "what" or "why". Users interact with AI agents with specific objectives in mind, and these outcome-related attributes are essential for fulfilling those objectives. Thus, these attributes are associated with the superordinate level of stimuli identification and reflect the AI agent's overall performance and effectiveness, which are likely to be evaluated better with a high construal frame. On the other hand, the process-related attributes can be considered peripheral attributes. These attributes describe "how" the user interacts with the AI agent, such as the ease of use, controllability, communication quality, etc. Although they contribute to the overall experience, they are secondary to the main goal of utilizing the AI agent. Hence, these aspects are linked to the subordinate level of stimuli identification and depict the operational details and procedures of the interaction, which would require a low construal frame to assimilate the information. Therefore, AI users are expected to assign relatively more importance to process-related attributes when evaluating an AI agent under a low-level construal and attribute greater significance to outcome-related attributes under a high-level construal.

RESEARCH MODEL AND PROPOSITIONS

Drawing on the Two-Factor Theory, CLT, and IS success and AI trust research, this paper proposes a research model of how user experience with AI agents influences users' trust and distrust in AI agents (Figure 1). It classifies the characteristics of user experience with AI agents into two categories: those related to the process and those related to the outcome. It suggests that these process-related and outcome-related attributes of the AI user experience have distinct effects on users' trusting and distrusting beliefs in AI agents. These effects can be moderated by the users' construal frames of evaluation and subsequently impact their continued intention to use AI agents. Moreover, the research model proposes that when there's a match— "construal fit"—between how users perceive AI agents and their construal frame of evaluation, it can enhance their trusting beliefs in AI agents.

Figure 1. Research Model of the Effects of User Experience Attributes on Trust, Distrust and Continuance Intention in AI agents



The Two Factor Theory suggests the asymmetric effects of process-related and outcome-related attributes of AI user experience on trusting and distrusting beliefs in AI agents. Process-related attributes, such as the ease of use, controllability, communication quality, sociability, and transparency of an AI agent, are factors that contribute to the interaction between the user and the AI agent. They are often seen as motivating factors that enhance users' positive attitudes towards and engagement with the AI agent (Vermeeren, Law, Roto, Obrist, Hoonhout & Väänänen-Vainio-Mattila, 2010). So, they primarily contribute to the development of trusting beliefs. The higher the quality of these attributes, the more likely users are to develop trust in the AI agent (Beldad, De Jong & Steehouder, 2010; Lee & See, 2004). For instance, if an AI agent is easy to use, transparent, and communicates clearly, it can foster a sense of confidence in the system, leading to increased trust. However, these process-related attributes might not have a considerable impact on reducing distrust, because even if the process is user-friendly, clear and smooth, users could still harbor distrust due to some outcome-related issues. Even if these attributes are of lower quality, they might not necessarily lead to strong distrusting beliefs, as users can be more tolerant of a suboptimal interactive process as long as the outcome meets their needs and expectations (Bhattacherjee, 2001).

On the other hand, the outcome-related attributes related to the AI agent's performance, such as usefulness, effectiveness, reliability, helpfulness, and efficiency, primarily serve to diminish distrusting beliefs rather than increase trusting beliefs (Herzberg, 1964), because these attributes are considered "must-haves" and act as hygiene factors, which, when absent or inadequate, can lead to negative attitudes, such as dissatisfaction or distrust (Herzberg, 1964; Kim, Ferrin & Rao, 2008). For example, if an AI agent makes errors or fails to complete tasks effectively, users may perceive the agent as unreliable or untrustworthy and start distrusting the system (Bhattacherjee, 2001; Gefen, Karahanna & Straub, 2003). If an AI agent consistently delivers accurate results and exhibits reliable performance, users' potential fears and doubts (distrusting beliefs) about the system are likely to decrease. However, even if the AI agent performs well, it does not necessarily mean that users will develop strong trusting beliefs, as good performance is often taken for granted (Oliver, 1980). Users may simply consider performance as an expected standard (Benbasat, Gefen & Pavlou, 2008) rather than a source of increased trust.

Therefore, the process-related attributes of AI user experience seem to play a more significant role in building trusting beliefs, while outcome-related attributes play a more critical role in mitigating distrusting beliefs. This suggests the following propositions.

> Proposition 1: Process-related attributes of AI user experience have an asymmetric effect on trusting beliefs and distrusting beliefs; they act more to increase trusting beliefs than to decrease distrusting beliefs in AI agents.
>
> Proposition 2: Outcome-related attributes of AI user experience have an asymmetric effect on trusting beliefs and distrusting beliefs, they act more to decrease distrusting beliefs than to increase trusting beliefs in AI agents.

CLT posits that psychological distance influences how people perceive, comprehend, and evaluate objects and events (Trope & Liberman, 2010). It suggests that individuals use two types of mental representations: high-level construals (abstract, schematic, and decontextualized representations) for distal stimuli, and low-level construals (detailed, specific, and contextualized representations) for proximal stimuli. In the context of AI user experience, it can be argued that process-related attributes are associated with low-level construals, while outcome-related attributes are linked with high-level construals. Process-related attributes, such as ease of use, controllability, communication quality, sociability, and transparency, are closely tied to the immediate interaction between users and AI agents, which users can evaluate in a detailed and specific manner. On the other hand, outcome-related attributes, such as usefulness, effectiveness, helpfulness, efficiency, and reliability, are often viewed in a higher-level and more abstract manner. So, users' construal levels can have important implications for how their process and respond to different aspects of AI agents when interacting with them. For instance, the construal frames of evaluation adopted by users of AI agents can affect the importance that users place on the process- or outcome-related attributes of AI user experience. Individuals adopting a high-level construal frame of evaluation often care more about the overall outcomes rather than the specific steps or processes involved in achieving those outcomes, and hence tend to place more emphasis on the outcome-related attributes, such as the performance of AI agents. Whereas individuals adopting a low-level construal frame of evaluation focus on the specific details of how to use AI agents to achieve tasks, and thus tend to place more emphasis on the process-related attributes, such as the ease of use, controllability, communication quality and so on.

In the context of AI agent usage, users may form different construal frames based on their usage contexts. For instance, if the AI interaction is immediate (temporally proximate), with a close friend (socially proximate), in the same room (spatially proximate), and certain to occur (high probability), it might induce a low-level construal frame of evaluation. For example, if you are using a smart speaker like Amazon's Alexa in your living room to play music or ask for the weather forecast, this situation is temporally, spatially, and socially proximate and certain to occur, leading to a low-level construal frame. In this scenario, you will focus more on the concrete and immediate details of the interaction process (such as Alexa's easiness of use, voice clarity, response speed,

and friendliness), which are characteristics of a low-level construal. On the contrary, if you are using an AI investment advisor to manage your retirement savings ten years from now, this situation is temporally distant, socially distant (as it's impersonal), and uncertain (as market conditions can vary), leading to a high-level construal frame. In this case, you might focus more on the abstract and ultimate outcomes of the interaction (e.g., the expected return on investment and the AI advisor's past performance), which are characteristics of a high-level construal.
In sum, the usage contexts of AI agents can shape users' construal frames of evaluation, which in turn influence their focus on the process- or outcome-related attributes of user experience and moderate the effects of these attributes on users' trusting and distrusting beliefs in AI agents.

Based on the above reasoning, the following propositions can be formulated.

> Proposition 3: Process-related attributes of AI user experience have a) a stronger positive influence on users' trusting beliefs and b) a stronger negative influence on users' distrusting beliefs in AI agents under a low-level construal frame than a high-level construal frame of evaluation.
>
> Proposition 4: Outcome-related attributes of AI user experience have a) a stronger negative influence on users' distrusting beliefs and b) a stronger positive influence on users' trusting beliefs in AI agents under a high-level construal frame than a low-level construal frame of evaluation.

Prior research has shown that people's mental representations of AI agents may be construed at different levels depending on people's beliefs about AI's learning capabilities (Kim & Duhachek, 2020). People tend to perceive AI agents as low-level construals when they hold a lay theory that AI agents do not have superordinate goals and cannot learn from their experiences or possess consciousness like humans do, because they are considered as fixed-capability machines that can follow only preprogrammed algorithms (Kim & Duhachek, 2020). In contrast, when individuals view AI agents as advanced, autonomous entities capable of complex problem-solving and learning via prior experiences, they are likely to infer greater superordinate goals from AI agents and consider them as high-level construals (Kim & Duhachek, 2020).

When a user's mental representation of an AI agent aligns with his/her construal frame of evaluation to achieve a construal fit, he/she is likely to process information from the AI agent more fluently. This is because the AI agent's behavior is in line with the user's expectation and cognitive framework, making it easier for him/her to understand and predict the AI agent's actions. Information that is easier to process tends to be perceived as more familiar, more pleasant, and more truthful (Reber, Winkielman & Schwarz, 1998; Winkielman, Schwarz, Fazendeiro & Reber, 2003). So, the processing fluency resulting from the construal fit between the users' perception of AI agents and their construal frame of evaluation will make the users feel more pleasant interacting with the AI agent and perceive the AI agent as more reliable, which in turn leads to increased trust. Furthermore, the construal fit induces a "feeling right" experience. This feeling of appropriateness can function as a source of information, as per the "feeling-as-information" theory (Schwarz & Clore, 2007), which suggests that people draw inferences from their feeling right experience, resulting in more favorable evaluations (Avnet & Higgins, 2006; Camacho, Higgins & Luger, 2003; Higgins, Idson, Freitas, Spiegel & Molden, 2003). In the context of AI usage, the "feeling right" experience due to a construal fit can generate a sense of trustworthiness about the AI agent because users interpret their "feeling right" as a signal that their interaction with the AI agent is going well and therefore, the agent can be trusted.

For example, when interacting with an AI agent, a user with a high-level construal frame of evaluation might focus on the agent's potential for learning, growth, adaptability, and its ability to provide personalized solutions. If this user perceives the AI system as a high-construal entity capable of learning and improving over time, they are more likely to trust the AI agent. This is because their high-level big-picture thinking aligns with the AI's perceived capabilities, creating a feeling of appropriateness and a sense of confidence in the AI's ability to perform tasks efficiently and effectively. On the other hand, a user with a low-level construal frame of evaluation focuses on concrete and specific aspects when evaluating a situation or an object. So, when interacting with an AI agent, this user might focus on the specific tasks the AI can perform and the immediate outcomes the AI provides. If this user perceives the AI agent as a low-construal machine with fixed capabilities to perform specific, repetitive tasks, he/she may trust the AI agent more because he/she feels right about relying on the agent that consistently delivers the expected results.

Therefore, the following propositions are suggested.

> Proposition 5: A construal fit between users' perception of AI agents and their construal frame of evaluation increases their trusting beliefs in AI agents.
> Proposition 5a: When users perceive AI agents as lower-level construals, they will have greater trusting beliefs in AI agents under a low-level construal frame than a high-level construal frame of evaluation.
> Proposition 5b: When users perceive AI agents as high-level construals, they will have greater trusting beliefs in AI agents under a high-level construal frame than a low-level construal frame of evaluation.

Trust and distrust play crucial roles in shaping users' attitudes toward AI agents and their intention to continue using them. Trusting beliefs in AI agents refer to the perception that the AI systems have positive attributes such as competence, integrity, and benevolence (McKnight, Choudhury & Kacmar, 2002). In contrast, distrusting beliefs denote negative attributes such as incompetence, deceit, and malevolence (Lewicki, McAllister & Bies, 1998).
Prior research has shown that trust positively impacts users' intention to continue using technology and increases user satisfaction (Bhattacherjee, 2001; Gefen, Karahanna & Straub, 2003). Trusting beliefs in AI agents can foster a sense of confidence and reliability in the system, leading users to continue using the AI agents for various tasks. For example, when users perceive an AI-driven recommendation system as trustworthy, they are more likely to rely on its suggestions and continue using the platform (Xu, Benbasat & Cenfetelli, 2013).

On the other hand, distrust negatively influences continuance intention and can lead to avoidance or abandonment of technology (Cho, 2006; Lewicki, McAllister & Bies, 1998). If users believe that AI agents possess negative attributes such as deceit or incompetence, they are less likely to continue using the AI agents, fearing potential harm, misinformation, or poor performance. For instance, distrust in AI medical diagnosis systems may prevent patients from accepting the system's recommendations, impacting their intention to continue using such systems (Abdul, Vermeulen, Wang, Lim & Kankanhalli, 2018). So, trusting beliefs in AI agents are likely to have a positive impact on users' continuance intention, while distrusting beliefs can lead to a negative influence on the intention to keep using AI agents. And the following propositions can be formulated:

Proposition 6: Trusting beliefs in AI agents will positively influence continuance intention to use AI agents.

Proposition 7: Distrusting beliefs in AI agents will negatively influence continuance intention to use AI agents.

DISCUSSIONS AND CONTRIBUTIONS

This paper proposes a research model of how user experience with AI agents influences users' trust and distrust in AI agents. The major contributions of this paper are as follows. First, this research draws on the Two-Factor Theory to differentiate between the process-related and outcome-related attributes of AI user experience and examine their differential effects on trusting and distrusting beliefs in AI Agents. Process-related attributes, classified as motivating factors, have stronger impacts on trusting beliefs, whereas outcome-related attributes, categorized as hygiene factors, have stronger impacts on distrusting beliefs. These asymmetric effects highlight the importance of both designing an interactive process that fosters trust and ensuring the AI agent's outcomes are reliable and effective to prevent distrust. Second, by applying CLT to AI user experience, this paper contributes to the understanding of the information-processing mechanism by which users' construal levels influence how AI users prioritize the process-related or outcome-related attributes of AI user experience in their evaluations of AI agents. It helps identify why some users may place more emphasis on the process of AI interaction (low-level construals) while others may care more about the overall outcomes of AI usage (high-level construals). Additionally, it also highlights the importance of construal fit between users' mental representations of AI agents and their construal frames of evaluation for increasing trust in AI agents. Finally, by providing an understanding of the role of construal fit in promoting trust and the psychological mechanism by which various attributes of AI user experience differentially influence users' trust and distrust in AI agents, this paper will offer guidelines on how to appropriately design and implement AI agents to enhance trust and minimize distrust. Incorporating these considerations into AI agent design can ensure that AI technologies are more widely accepted and adopted across various industries and applications. For example, an AI system designed for a user who is process-oriented should focus more on making it easy to use and control as well as transparent and understandable. In contrast, an AI system designed for a user who is outcome-oriented should focus more on ensuring its outcomes are accurate, useful, and reliable.

REFERENCES

Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-18). ACM.

Avnet, T., & Higgins, E. T. (2006). How regulatory fit affects value in consumer choices and opinions. *Journal of Marketing Research*, *43*, 1-10.

Bansal, R., Pruthi, N., & Singh, R. (2022). Developing customer engagement through artificial intelligence tools: Roles and challenges. In J. Kaur, P. Jindal, and A. Singh (Eds.), *Developing relationships, personalization, and data herald in marketing 5.0* (pp. 130-145). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-6684-4496-2.ch008

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, 35(4), 1017-1041.

Beldad, A., de Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior*, *26*(5), 857-869.

Benbasat, I., Gefen, D., & Pavlou, P. A. (2008). Special issue: Trust in online environments. *Journal of Management Information Systems*, *24*(4), 5-11.

Bhattacherjee, A. (2001). Understanding information systems continuance: an expectation-confirmation model. *MIS Quarterly*, *25*(3), 351-370.

Cassell, J. (2000). Embodied conversational interface agents. *Communications of the ACM*, *43*(4), 70-78.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340.

Camacho, C. J., Higgins, E. T., & Luger, L. (2003). Moral value transfer from regulatory fit: What feels right is right and what feels wrong is wrong. *Journal of Personality and Social Psychology*, *84*, 498-510.

Cho, J. (2006). The mechanism of trust and distrust formation and their relational outcomes. *Journal of Retailing*, *82*(1), 25–35.

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, 114–126.

Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*. 261-268.

Freitas, A. L., Salovey, P., & Liberman, N. (2001). Abstract and concrete self-evaluative goals. *Journal of Personality and Social Psychology, 80*(3), 410–424.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254-280.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, *27*(1), 51-90.

Griffin, D. W., & Ross, L. D. (1991). Subjective construal, social inference, and human misunderstanding. *Advances in Experimental Social Psychology, 24*, 319-359.

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Hassen, A. B. H., Thomas, L., Enk, A., Uhlmann, L., Reader study level-I and level-II Groups, Alt, C., Arenbergerova, M., Bakos, R., Baltzer, A., Bertlich, I., Blum, A., Bokor-Billmann, T., Bowling, J., … Zalaudek, I. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology: official journal of the European Society for Medical Oncology*, *29*(8), 1836–1842.

Hassenzahl, M., & Tractinsky, N. (2006). User experience - a research agenda. *Behaviour & Information Technology*, *25*(2), 91-97.

Henry, S. L., Abou-Zahra, S., & Brewer, J. (2014). The role of accessibility in a universal web. In *Proceedings of the 11th Web for All Conference* (pp. 1-4).

Herzberg, F. (1964). The motivation-hygiene concept and problems of manpower. *Personnel Administration*, *27*(1), 3-7.

Herzberg, F., (1968). One more time: how do you motivate employees? *Harvard Business Rev*iew, *46*(1), 53–62.

Herzberg, F., Mausner, B., & Snyderman, B. B. (1959). *The motivation to work*. New York: John Wiley & Sons.

Higgins, T. E., Idson, L. C., Freitas, A. L., Spiegel, S., & Molden, D. C. (2003), Transfer of value from fit. *Journal of Personality and Social Psychology*, *84*, 1140-1153.

Higgins, E. T., & Trope, Y. (1990). Activity engagement theory: Implications of multiply identifiable input for intrinsic motivation. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 229-264). New York: Guilford Press.

Jackson, P. C. (2019). *Introduction to artificial intelligence*. New York: Courier Dover Publications.

Jiang, Z., Benbasat, I., & Wang, W. (2018). Maintaining and enhancing the quality of communication between humans and AI. *MIS Quarterly Executive*, *17*(2), 71-83.

Johnston, R. (1995). The determinants of service quality: Satisfiers and dissatisfiers. *International Journal of Service Industry Management*, *6*(5), 53–71.

Kaplan, A. M., & Haenlein, M. (2019) Siri, siri, in my hand: who's the fairest in the land? On the interpretations, illustrations and implications of artificial intelligence. *Business Horizons, 62*(1), 15–25. https://doi.org/10. 1016/j.bushor.2018.08.004 29.

Kim, T., & Duhachek, A. (2020). Artificial intelligence and persuasion: A construal-level account. *Psychological Science*, *31*. 095679762090498. 10.1177/0956797620904985.

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems*, *44*(2), 544-564.

Knijnenburg, B. P., Willemsen, M. C., Ganter, Z., Soncu, H., & Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, *22*, 441-504.

Komiak, S. Y. X., & Benbasat, I. (2008). A two-process view of trust and distrust building in recommendation agents: a process-tracing study. *Journal of the Association for Information Systems*, *9*(12), 727–747.

Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. *Annual Review of Psychology*, *50*, 569–598.

Kruglanski, A. W. (1975). The endogenous-exogenous partition in attribution theory. *Psychological Review*, *82*(6), 387-406.

Lankton, N. K., Wilson, E. V., & Mao, E. (2010). Antecedents and determinants of information technology habit. *Information & Management*, *47*(5-6), 300-307.

Law, E. L. C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009). Understanding, scoping and defining user experience: A survey approach. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 719-728.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80.

Lewicki, R. J., McAllister, D. J., & Bies, R. J. (1998). Trust and distrust: new relationships and realities. *The Academy of Management Review*, *23*(3), 438–458.

Liberman, N., & Trope, Y. (2008). The psychology of transcending the here and now. *Science*, *322*(5905), 1201–1205.

Liberman, N., Trope, Y., & Wakslak, C. (2007). Construal level theory and consumer behavior. *Journal of Consumer Psychology*, *17*(2), 113-117.

Liberman, N., Sagristano, M. D., & Trope, Y. (2002). The effect of temporal distance on level of mental construal. *Journal of Experimental Social Psychology*, *38*(6), 523-534.

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology, 75*(1), 5–18.

Loiacono, E. T., Watson, R. T., & Goodhue, D. L. (2007). WebQual: an instrument for consumer evaluation of web sites. *Journal of Electronic Commerce*, *11*(3), 51–87.

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.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: an integrative typology. *Information Systems Research*, *3*(3), 334–359.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*(2), 246–268.

Moody, G. D., Lowry, P. B., & Galletta, D. F. (2017). It's complicated: Explaining the relationship between trust, distrust, and ambivalence in online transaction relationships using polynomial regression analysis and response surface analysis. *European Journal of Information Systems, 26*(4), 379-413.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98-100.

Novak, T. P., Hoffman, D. L., & Yung, Y. F. (2000). Measuring the customer experience in online environments: A structural modeling approach. *Marketing Science*, *19*(1), 22-42.

Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, *17*(4), 460-469.

Ou, C. X., & Sia, C. L. (2009). To trust or to distrust, that is the question: investigating the trust-distrust paradox. *Communications of the ACM*, *52*(5), 135–139.

Ou, C. X., & Sia, C. L. (2010). Consumer trust and distrust: an issue of website design. *International Journal of Human Computer Studies*, *68*(12), 913–934.

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL: A multiple-item scale for assessing electronic service quality. *Journal of Service Research*, *7*(3), 213-233.

Pavlou, P. A., & Gefen, D. (2004). Building effective online marketplaces with institution-based trust. *Information Systems Research*, *15*(1), 37-59.

Picard, R. W. (1997). *Affective computing*. Cambridge: MIT Press.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347-1358.

Reber, R., Winkielman, P., & Schwarz, N. (1998). Effects of perceptual fluency on affective judgments. *Psychological Science*, *9*(1), 45–48.

Russell S., & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Upper Saddle River: Prentice Hall.

Schaefer, K. E., Chen, J. Y., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human factors*, *58*(3), 377–400.

Schwarz, N., & Clore, G. L. (2007). Feelings and phenomenal experiences. *Social Psychology: Handbook of Basic Principles*, *2*, 385–407.

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, *31*, 47-53.

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic “remedies” for trust/distrust. *Organization Science*, *4*(3), 367–392.

Szolovits, P. (ed.) (2019). *Artificial intelligence in medicine*. London: Routledge.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review, 110*(3), 403–421.

Trope, Y., & Liberman, N. (2000). Temporal construal and time-dependent changes in preference. *Journal of Personality and Social Psychology, 79*(6), 876–889.

Trope, Y., Liberman, N., & Wakslak, C. (2007). Construal levels and psychological distance: Effects on representation, prediction, evaluation, and behavior. *Journal of Consumer Psychology*, *17*(2), 83-95.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.

Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57*(4), 660–671.

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*(1), 3-15.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.

Verhagen, T., van Nes, J., Feldberg, F., & van Dolen, W. (2014). Virtual customer service agents: Using social presence and personalization to shape online service encounters. *Journal of Computer-Mediated Communication*, *19*(3), 529-545.

Vermeeren, A. P., Law, E. L., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: current state and development needs. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries* (pp. 521-530). ACM.

Wiesenfeld, B. M., Reyt, J. N., Brockner, J., & Trope, Y. (2017). Construal level theory in organizational research. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 367-400.

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189-217). Erlbaum.

Xingyuan, W., Li, F., & Wei, Y. (2010). How do they really help? An empirical study of the role of different information sources in building brand trust. *Journal of Global Marketing*, *23*(3), 243-252.

Xu, H., Benbasat, I., & Cenfetelli, R. T. (2013). Integrating service quality with system and information quality: An empirical test in the e-service context. *MIS Quarterly*, *37*(3), 777-794.

QRBD

Quarterly Review of Business Disciplines

February 2024

Volume 10
Number 3/4

A Journal of International Academy of Business Disciplines
Sponsored by University of North Florida
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)