Computers in Human Behavior 115 (2021) 106607

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

Attachment and trust in artificial intelligence

Omri Gillath [a,*], Ting Ai [a], Michael S. Branicky [b], Shawn Keshmiri [c], Robert B. Davison [d], Ryan Spaulding [e]

[a] Department of Psychology, University of Kansas, 1415 Jayhawk Blvd., Lawrence, KS, 66045-7556, USA
[b] Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, 66045, USA
[c] Department of Aerospace Engineering, University of Kansas, Lawrence, KS, 66045, USA
[d] Department of Management, College of Business, Colorado State University, CO, 80523, USA
[e] Department of Biostatistics & Data Science, School of Medicine, University of Kansas, Kansas City, KS, 66160, USA

ARTICLE INFO

Keywords:
Artificial intelligence
Attachment style
Close relationships
Trust

ABSTRACT

Lack of trust is one of the main obstacles standing in the way of taking full advantage of the benefits artificial intelligence (AI) has to offer. Most research on trust in AI focuses on cognitive ways to boost trust. Here, instead, we focus on boosting trust in AI via affective means. Specifically, we tested and found associations between one's attachment style—an individual difference representing the way people feel, think, and behave in relationships—and trust in AI. In Study 1 we found that attachment anxiety predicted less trust. In Study 2, we found that enhancing attachment anxiety reduced trust, whereas enhancing attachment security increased trust in AI. In Study 3, we found that exposure to attachment security cues (but not positive affect cues) resulted in increased trust as compared with exposure to neutral cues. Overall, our findings demonstrate an association between attachment security and trust in AI, and support the ability to increase trust in AI via attachment security priming.

Artificial intelligence (AI) is a broad field encompassing computer systems that can complete tasks normally requiring human intelligence, such as visual perception, speech recognition, and decision-making under uncertainty (Russell & Norvig, 2020; Rossi, 2018). The definition of AI has evolved over time (Brachman, 2006), but fundamentally also includes adaptation and learning from experience (Wang, 2008). Herein, we adopt the following working definition, which was used in our study questionnaires:

Artificial Intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions), and self-correction. Particular applications of AI include expert systems, speech recognition, and machine vision. Examples of AIs include personal helpers (like Siri and Alexa), medical diagnostic aids, and self-driving vehicles.

Although the prevalence of and the roles AI plays have increased exponentially (e.g., Stormont, 2008; You & Robert, 2018), many people still have a hard time trusting AI (Andriole, 2018; Bloomberg, 2018; Chang, 2016; Pegasystems, 2018; Towers-Clark, 2019). For example, in one survey on whether people trust AI, only 44% of participants said that if a self-driving Uber car picked them up, they would get in (DriversEd.com, 2019). A more recent survey by SurveyUSA (2020) of 1200 adults showed similar results, such that 48% of the people said they would never get in a self-driving vehicle, 21% said they were unsure, and 20% said that autonomous vehicles would never be safe. Other surveys reveal that the fear is more general, for instance, 42% of people lack general trust in AI, and 49% of people could not name a single AI product they trusted (Dujmovic, 2017).

There might be different reasons why people do not trust AI. Some people do not understand AI, hence the field of research known as explainable AI, or XAI, which attempts to build human-comprehensible and interpretable systems that can explain their decisions (Doran, Tamma, & Iannone, 2007; Samek, Montavon, Vedaldi, Hansen, & Müller, 2019). Others are afraid of AI (Andriole, 2018). For example, people might be scared of losing their jobs because of AI, or even lose their lives. Indeed, a recent survey conducted by Oxford University's Center for the Governance of AI, showed that many Americans fear AI (Zhang & Dafoe, 2019). Americans ranked a possible AI apocalypse as more catastrophic than a possible failure to address climate change. In

* Corresponding author.
E-mail address: ogillath@ku.edu (O. Gillath).

https://doi.org/10.1016/j.chb.2020.106607
Received 3 April 2020; Received in revised form 18 September 2020; Accepted 18 October 2020
Available online 21 October 2020
0747-5632/

line with these results, Liang and Lee (2017) using a nationally representative sample showed that over 25% of the people experience heightened fear of autonomous robots and artificial intelligence (FARAI). They further showed that this fear was positively correlated with other fears such as fear of drones or of becoming unemployed (see also Berg et al., 2018).

Regardless of the reason, lack of trust can result in reduced cooperation, efficiency, and productivity (e.g., Braynov & Sandholm, 2002; Siau & Wang, 2018). It can also reduce or prevent the integration of AI systems and agents into teams (Chakraborti et al., 2017; Groom & Nass, 2007), and the adoption of new technologies more broadly (Jeffries & Reed, 2000). According to McKnight and colleagues (McKnight et al., 2002) to create initial trust, perceptions of risk must be overcome, which in turn will create the willingness to use the technologies. For the casual user (and many people are casual users), the affective aspect of trust might be more impactful than the cognitive aspect. In other words, reducing fear and overcoming risk perceptions is likely to be more efficient, especially at the initial stages of trust, than providing information about number of errors or similar cognitive aspects. The goal of the current paper is to increase the understandings of the underlying mechanisms of trust in AI and based on these mechanisms identify new ways to increase trust via its affective aspect (e.g., reducing anxiety).

Two main routes are mentioned in the literature to boost trust in AI: a cognitive route and an affective one (Cook & Wall, 1980; Johnson & Grayson, 2005). The majority of research on trust in non-humans (i.e., autonomous machines, robots, and AIs) is focused on the cognitive route to trust (e.g., Hancock et al., 2011; Stormont, 2008). Relatively less work was done on the affective route and on psychological factors that could enhance trust via that route. The current studies were set to fill this gap in the literature by focusing on affective factors that could be used to increase overall trust in AI. One such potential factor is people's *attachment style* (Bowlby, 1982; for a review see; Gillath et al., 2016).

Attachment style, the way people feel, think, and behave in relationships, is often assessed as attachment security or insecurity (levels of anxiety and avoidance; Ainsworth et al., 1978). Numerous studies have shown attachment style to be an efficient predictor of various relational outcomes (for reviews see Cassidy & Shaver, 2018; Gillath et al., 2016). Specifically, a few studies on human relations showed that attachment security is associated with more trust, whereas attachment insecurity is associated with less trust in other humans (Mikulincer, 1998; Simmons, Gooty, Nelson, & Little, 2009).

Increasing trust in AI can increase productivity over time, performance, and human-AI team effectiveness, while reducing stress and various risks (e.g., Chakraborti et al., 2017; Thau et al., 2007). Here we tested whether attachment style predicts trust in AI and whether enhancing people's sense of attachment security can increase overall trust in AI.

Interpersonal Trust. As noted above, there are two routes to increase trust: (1) *affective* or emotional-based trust boost, and (2) *cognitive trust* boost or competency-based boost (Cook & Wall, 1980; Johnson & Grayson, 2005). The affective route to boost trust is defined as an increase in the faith in the trustworthy intentions of others, or the confidence people place in others based on how they feel about them (e.g., Bliss et al., 1994). These feelings can be generated by the level of care and concern others demonstrate (Johnson-George & Swap, 1982; Rempel et al., 1985). In other words, affective trust is based on a sense of security people gain from emotional bonds with others (Johnson & Grayson, 2005; Lewis & Weigert, 1985). This emotion-driven boost of trust is more subjective and less transparent to outsiders or procedures like risk assessment (Johnson & Grayson, 2005). It is also thought to be associated with actions that are intrinsically (as opposed to extrinsically) motivated (Rempel et al., 1985).

Cognitive boost of trust is related to people's confidence in- or willingness to rely on-others and their ability. It involves attributions of capability, competence, and reliability (Moorman et al., 1992; Rempel et al., 1985). The cognitive route is knowledge-driven, arising from accumulative knowledge that allows people to predict whether others will live up to their obligations or not (Johnson & Grayson, 2005; Moorman et al., 1992). People are more likely to trust others who make fewer mistakes (Akash et al., 2017, pp. 1542–1548; Jiang et al., 2004), others who are experts, and others or products that performed better in the past. Of central importance to our research, people tend to expect perfection from non-humans machines, so trust is severely decreased when errors occur (Dzindolet et al., 2002). The affective and cognitive routes to boost trust are thought to be correlated, yet empirically distinguishable—a feeling vs. a judgment. The affective route is more subjective and less transparent (e.g., harder for economists to prescribe risk assessments). It is also correlated with the view of partners' actions as intrinsically rather than extrinsically motivated (Rempel et al., 1985). Johnson and Grayson (2005) argue that the outcome (overall trust, or the way people behave) is based on both the cognitive and the affective routes. However, as emotional bonds deepen and security increases, trust increases, venturing beyond the trust justified merely by cognitive components. In other words, affect and security can go beyond what people get from the cognitive route to increase trusting behavior. Indeed, Webber (2008) showed that affective trust has a stronger positive relationship with team performance than cognitive trust.

Trust is important in interpersonal relationships, as well as in interpersonal dyadic negotiations (Lu et al., 2017), in providers—users relations (Moorman et al., 1992), and, of importance to the current paper, in human relations with non-humans (e.g., Hoff & Bashir, 2015; Parasuraman & Riley, 1997). Here we will extend these lines of work to trust in AI.

Trust in non-humans. Interpersonal trust models play a key role in understanding human interaction with non-human entities (like automation). Studies have found that human-automation interaction adopts similar norms to human-human interaction (Madhavan & Wiegmann, 2007). This allows for the application of interpersonal trust theories to human-automation relationships (but see Evers et al., 2008; Groom & Nass, 2007; Sanders et al., 2011). Research shows that cognitive factors related to the robot like performance and descriptive attributes (serious vs. pet-like) associate with trust in non-humans (Hancock et al., 2011; Robinson, MacDonald, & Broadbent, 2014). Some of these factors are thought to make machines or robots more human-like, and in turn increase cognitive trust in non-humans (Waytz et al., 2014; You & Robert, 2019).

Indeed, automation characteristics that increase anthropomorphism have been found to increase trust as well (Crandall et al., 2018; Waytz et al., 2014). This link can be attributed to the perception of competence in the automation when it looks human-like (Waytz et al., 2014). AI systems in humanoid form or even pet-like forms induce more trust, whereas aggressive-looking systems hinder trust (Siau & Wang, 2018). Non-humans that share goals and provide information about their actions are perceived as more trustworthy by people (Verberne et al., 2012). For example, a shared policy of medical treatments increased trust in the AI equivalent to humans (Yokoi & Nakayachi, 2018).

A relatively small number of studies have examined affective boosts for trust and the psychological factors that affect trust in non-human (Hancock et al., 2011). For example, high conscientiousness was found to predict dislike for a social robot that uses a physical interface (because conscientious people preferred a text interface; Looije et al., 2010). In a different study, Sarkar et al. (2017) found that people high on extraversion perceived interactions with robots more positively, and trusting the robots was a part of that overall positivity. Mou and Xu (2017) have demonstrated that when interacting with AI, people showed less openness, agreeableness, extroversion, conscientiousness, and self-disclosure than when interacting with humans. Lower levels of these tendencies might explain why people are less likely to trust AIs.

Similar results were found with regard to trust in AI. For example, positive affect was shown to be positively correlated with trust in AI (Hughes et al., 2009). Conversely, past negative experiences were shown to be associated with decreased trust in AI (Dikmen & Burns, 2017), in a fashion similar to trust in automation (Dikmen & Burns, 2017; Hengstler et al., 2016). Negative attitudes towards technology were also found to predict low trust in AI (self-driving taxis; Tussyadiah et al., 2017). One personality factor that is less studied when it comes to trust in AI is attachment style.

Attachment and trust in AI. Attachment style is an individual difference known to play a significant role in human relations and their emotional bonds (Gillath et al., 2016). It develops over time based on the interactions people have with their primary caregivers. Sensitive responsive caring results in a secure attachment style, intrusive inconsistent caring results in anxious attachment style (high attachment anxiety), and cold rejecting insensitive caring results in avoidant attachment style (high attachment avoidance). People high on attachment anxiety, are preoccupied with thoughts about rejection and abandonment. They are often overwhelmed by these thoughts and the lack of ability to regulate their emotions, and constantly want to get closer to relationship partners. People high on attachment avoidance circumvent closeness and intimacy and do not want to depend on others or have others depend on them. People who are low on both anxiety and avoidance (securely attached), have long-lasting satisfying relationships, and do not experience any issues with being too close or not close enough to others.

Hundreds of studies demonstrate that people's attachment style is associated with various relational variables and affect regulation (for reviews see Cassidy & Shaver, 2018; Gillath et al., 2016). Attachment security is positively correlated with ease of forming and maintaining close relationships, and with higher commitment, intimacy, love, and satisfaction in such relationships; whereas attachment insecurity is associated with lower levels of these variables. Relevant to the current paper, very few studies have examined the associations between attachment and trust, showing that among humans, attachment security is associated with higher levels of trust, whereas attachment insecurity is associated with lower levels of trust (Mikulincer, 1998; Pistole, 1993; Simmons et al., 2009).

Recent findings suggest that attachment might play a similarly important role in human-AI interactions. For example, Birnbaum and colleagues showed that humans desire a robot's presence under stressful circumstances in a similar manner to their proximity-seeking behavior toward humans (termed attachment figures; Birnbaum et al., 2016). According to Baumeister and Leary (1995), humans crave companionship—they have an inherent need to belong and be a part of a relationship or an emotional tie (Ryan & Deci, 2000). This need, which can be fulfilled via a relationship with a human, such as a family member, friend, or co-worker, can potentially also be fulfilled by an inanimate object such as a robot or AI (Birnbaum et al., 2016; Siau & Wang, 2018). People may project human-like qualities such as autonomy, intelligence, and dependency onto robots, which in turn allow them to create an attachment bond with the robot/machine, and fulfill their need to belong or gain attachment security through that bond (see also Darling, 2016). Morsünbül (2018) provided further support for this idea by showing that children create strong emotional ties with caregiving robots. Perhaps this is because the robots can be taught to be responsive (Birnbaum et al., 2016) or due to anthropomorphism (Sugiyama & Vincent, 2013) as both have been found to contribute to the formation of attachment bonds.

Other studies suggest that factors related to secure attachment such as commitment and faithfulness or sociability and bonding (Pantic et al., 2007, pp. 47–71; Siau & Wang, 2018) increase the chances of creating an emotional bond with robots/AI. Thus, a higher affective commitment was found to positively correlate with creating an emotional bond with a robot (You & Robert, 2018), and creating an attachment with a dog-like robot was correlated with perceiving the robodog as faithful (Konok et al., 2018). Finally, the adoption of robots or AI was more likely to happen if attachment behaviors (neediness) were expressed by the robot/AI (Hiolle et al., 2012).

Here we will test whether attachment style can predict trust in AI. We expect attachment insecurity to be associated with less trust in AI. We will also test whether increasing attachment insecurity will decrease trust, whereas increasing attachment security will increase trust in AI. We predict that exposing people to attachment anxiety-related cues will decrease trust in AI; whereas exposure to security-related cues will increase trust.

1. Study 1

The purpose of Study 1 was to examine the associations between attachment style (levels of attachment anxiety and avoidance) and trust in AI, using a new measure we created. Attachment insecurity is known to be associated with a lack of trust among humans (Mikulincer, 1998). We, therefore, expected attachment insecurity to also predict trust (or lack of) between humans and AI. Because attachment insecurity, especially anxiety, is often associated with neuroticism (Noftle & Shaver, 2006) and self-esteem (Hart et al., 2005), we wanted to make sure that the associations between attachment anxiety and trust in AI were not due to these more general personality traits. Therefore, we also measured neuroticism and self-esteem. We predicted that attachment insecurity will be negatively associated with trust in AI, even after controlling for participants' neuroticism and self-esteem.

1.1. Method

Participants. A priori power analysis using the G*Power 3 computer program (Faul et al., 2007) indicated that a minimum sample of 68 people would be needed to detect a medium effect size ($d = 0.15$) with 80% power using an F-test with alpha at .05. Two-hundred and forty-eight participants were recruited using the snowballing technique, whereby research assistants use their social networks to provide potential participants a link to the online battery, and the participants then pass the link forward to other potential participants and so on. Participants volunteered to complete the survey without any compensation. Participants were mostly females (178 females, 70 males) and White (88.4% White, 6.8% Latino/Hispanic, 1.2% Black, 0.4% Asian or Pacific Island, and 3.2% multi-ethnic) with a wide age range (range 18–80, median = 21.5 years).

Materials and Procedure. After consenting, participants completed an online battery of self-report measures. The measures were divided into blocks, the order of the first two blocks was counterbalanced. One block included all the AI-related scales, another block included measures of adult attachment, big-five personality traits, and self-esteem. The third block included demographics (this block always came last).

Adult attachment style. Adult attachment style was measured using a short version of the Experiences in Close Relationships scale (ECR-16, Lo et al., 2009). ECR-16 consists of two subscales. One subscale assesses attachment anxiety (eight items; e.g., "I need a lot of reassurance that I am loved by people with whom I feel close," Cronbach $\alpha = 0.87$) and the other subscale assesses avoidant attachment (eight items; e.g., "I try to avoid getting too close to other people," $\alpha = 0.80$). Participants were asked to think about their close relationships, without focusing on a particular one or a specific partner, and rate the extent to which each item accurately described their feelings and behavior in these relationships using a 7-point response scale (ranging from 1 = *not at all*, to 7 = *very much*). Two scores were computed by averaging items on each subscale after appropriately reverse-scoring some of the items. The two subscales were slightly positively correlated, $r = 0.14$, $p < .05$.

Neuroticism. Neuroticism was measured with the Ten-Item Personality Inventory (TIPI, Gosling et al., 2003). The TIPI includes 10 items measuring five broad personality traits: extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience. The items consist of short phrases (e.g., anxious, easily upset) related to each trait. Participants indicated the extent to which each phrase applies to them on a 7-point scale (1 = *disagree strongly*, 7 = *agree strongly*). In the present study, the alpha for the neuroticism scale, the only one used in our analyses was α = 0.63.

Self-esteem. To assess self-esteem, we used the Rosenberg Self-Esteem Scale (Rosenberg, 1965). The scale consists of 10 items that are rated on a 7-point scale from 1 "*strongly disagree*" to 7 "*strongly agree*". Sample items include "*On the whole, I am satisfied with myself.*" The coefficient alpha of the measure was high, $\alpha = 0.90$.

Experience with AI. As the experience with AI and knowledge about AI might vary between participants, and might be associated with their trust in AI, we measured participants' previous experience with AI. We first provided a brief definition of AI:

"Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions), and self-correction. Particular applications of AI include expert systems, speech recognition, and machine vision. Examples of AIs include personal helpers (like Siri and Alexa), medical diagnostic aids, and self-driving vehicles."

Participants then reported whether they are using any AI-based technology or not and what technology they are using. Results showed that about 37% of participants reported they were not using any AI-based technology. For those who were using AI-based technology, Google, Siri, and Alexa were the most frequently mentioned.

Familiarity with AI. Familiarity with AI was measured with one item "How familiar are you with AI?" using a 5-point Likert-type scale (1 = extremely familiar, 5 = not familiar at all). This item was then reverse-scored to make the interpretation of the results easier. Higher scores mean being more familiar with AI. Results showed that overall participants were slightly familiar with AI (Mean = 2.62, SD = 0.88).

Trust in AI. We created six application scenarios describing potential interactions people have with AI in their daily lives, including self-driving vehicles/autopilot, medical diagnostic aids, and personal relationship aids. Each application was measured with two scenarios. Shown below are scenarios from each application.

Self-driving vehicles: You are scheduling an Uber ride via your online app. When the car gets there to pick you up, you find out it is a self-driving car.

You are buying a flight ticket to go on vacation. When you search the available flights, you find a flight that matches exactly your expected travel dates, destination, and price. Upon closer look, you learn that the flight will be with a self-flying airplane (autopilot).

Medical diagnostic aids: You are developing some symptoms and do not feel well. It is late at night and you do not want to go to the emergency room, you decide instead to use an online app to diagnose your illness. When you log in to the app you find out that an AI is making the clinical diagnosis.

You go to the hospital because you are experiencing pains. When you get there, you see that an AI is making the clinical diagnosis for patients.

Personal relationship aids: You are using a dating app to meet someone. After you input your detailed information, you find out that an AI will select a potential mate for you.

You feel lonely because you have few friends. You want to hang out with someone but always have no idea whom you could call. One day, you hear about an AI, which is made as a companion for humans, sells well, and you could afford the price.

Participants were instructed to read each scenario and then using a 7-point Likert-type scale (1 = not at all, 7 = extremely) to indicate: 1) How likely they are to take the car/accept their AI diagnosis/follow their dating AI suggestions/acquire an AI as a companion; 2) How likely they are to trust the AI; 3) How much they feel physically safe to ride this car/be diagnosed by the AI/follow the AI's suggestions/have an AI as a companion; 4) How emotionally secure do they feel to ride this car/be diagnosed by the AI/follow their AI's suggestions/have an AI as a companion. We used an exploratory factor analysis to examine whether the four items represented one factor (trust), two (affective trust vs. cognitive trust) or more (scores were averaged across the six scenarios for each item). Results showed a strong single factor that explained 92.52% of the variances, with all items loading higher than 0.90 on this factor. Therefore, scores were averaged across the six scenarios, and then across the four applications to compute the *Trust in AI* total score. Cronbach alpha for the measure was high, $\alpha = 0.96$.

Table 1
Descriptive statistics and correlations.

Variable	Mean (SD)	1	2	3	4	5	6
Study 1							
1. Trust	4.02 (1.20)						
2. Log_Age	1.41 (.16)	-.12					
3. Familarity	2.63 (.89)	.16*	-.06				
4. Attachment anxiety	3.98 (1.30)	-.08	-.38**	.04			
5. Attachment avoidance	3.01 (1.03)	.05	-.11	.02	.14*		
6. Self-esteem	3.48 (1.15)	-.04	.35**	.10	-.41**	-.36**	
7. Neuroticism	4.35 (1.45)	.02	.31**	.05	-.49**	-.11	-.52**
Study 2							
1. Trust	3.10 (1.34)						
2. Log_Age	1.47 (.16)	-.16**					
3. Familarity	2.90 (1.00)	.25**	-.002				
Study 3							
1. Trust	2.79 (1.13)						
2. Log_Age	1.48 (.19)	-.21**					
3. Familarity	2.66 (.96)	.36**	-.07				
4. Attachment anxiety	3.91 (1.29)	.07	-.41**	-.09			
5. Attachment avoidance	3.11 (1.00)	.08	-.11	-.08	.13*		

1.2. Results and discussion

Descriptive statistics and correlations among trust and all individual difference variables are presented in Table 1. To test our hypotheses regarding the association between attachment style and trust in AI, we conducted a hierarchical regression analysis predicting trust in AI (see Table 2; for unstandardized regression coefficients for all predictors.). In step 1, we entered age (to control for potential age effects; Knowles & Hanson, 2018) and familiarity with AI. The distribution of age was positively skewed, with most of the people falling between 18 and 40, with a few outliers above 40 (see Fig. 1). To resolve that skewness, we

Table 2
Regression coefficients for trust in AI as a function of age, attachment anxiety, attachment avoidance, self-esteem, and neuroticism. Study 1.

Predicting variables	Step 1	Step 2	Step 3
Intercept	3.722	3.853	4.102
Age	-.011	-.015*	-.013*
Familiarity	.235*	.230*	.245*
Attachment avoidance		.082	.069
Attachment anxiety		-.178*	-.220*
Self-esteem			-.044
Neuroticism			-.041
R^2	.045	.066	.070
R^2 change	.045	.020	.004

Note: All coefficients represent unstandardized regression coefficients. *$p < .05$. **$p < .001$. n = 232.

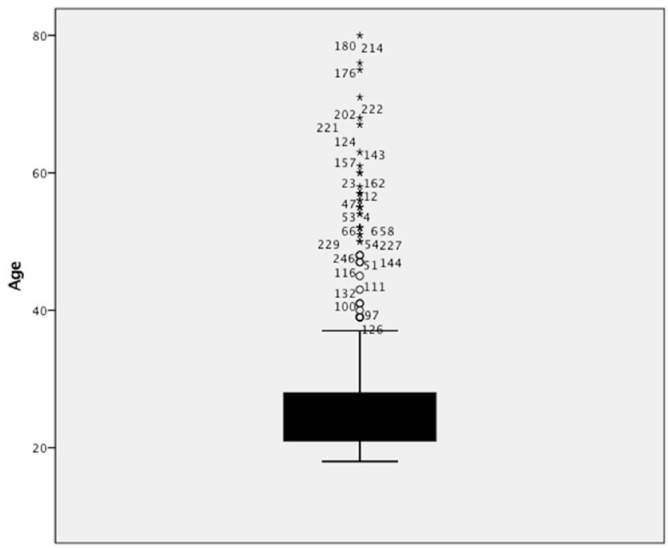

Fig. 1. The boxplot for age –Study 1.

used a log transformation of the age in the analyses. In step 2, we entered standardized attachment avoidance and anxiety scores. And in step 3 we entered neuroticism and self-esteem.

We did not find a significant main effect of age. However, we did find a significant main effect of familiarity with AI, such that as the familiarity with AI increased, participants were more likely to trust AI. We also found a significant main effect of attachment anxiety, consistent with our predictions, such that the higher one's anxiety score was, the less likely s/he was to trust AI. We did not find any effect of avoidant attachment, self-esteem, or neuroticism. Moreover, the main effect of attachment anxiety remained significant after controlling for self-esteem and neuroticism. This means the association between attachment anxiety and trust in AI could not be attributed to age, familiarity with AI, or broader personality traits like self-esteem and neuroticism.

2. Study 2

In Study 1, we found that the higher one's attachment anxiety was the lower her reported trust in AI was. However, we were unable to infer directionality and causality for this link, as our study design was correlational. Therefore, it is unclear whether attachment style leads to changes in trust or changes in trust result in changes to attachment (feeling more (in)secure). As we are interested in the possibility to change trust in AI, figuring out directionality and causality are highly important. Therefore, in Study 2, we switched to an experimental design and used well-studied attachment-related primes (e.g., Bartz & Lydon, 2004). We randomly assigned participants to one of three priming conditions (attachment anxiety, avoidance, or security). We then asked participants to recall a relationship that activated an attachment model in line with their condition and to write about that relationship for a few minutes. This method was successfully used by us and others multiple times in the past to activate attachment-related models (e.g., Gillath et al., 2010; for a review see; Gillath & Karantzas, 2019). We predicted that priming attachment anxiety will lead to a decrease in reported trust as compared with priming attachment security. In light of the lack of results for avoidance in Study 1, we did not expect attachment avoidance priming to have a similar effect.

2.1. Method

Participants. We used a between-subjects design with three conditions. A priori power analysis using the G*Power 3 computer program (Faul et al., 2007) indicated that a total sample of at least 159 people would be needed to detect a medium effect size ($d = 0.15$) with 80% power using an F-test with alpha at .05. We ended up recruiting a total sample of 374 participants. Two hundred and five of the participants were recruited in the same way as in Study 1 (snowballing technique), 169 were recruited at the crowdsourcing platform, Amazon Mechanical Turk (MTurk) with $1.00 as the compensation. Participants were mostly females (230 females, 142 males, 2 did not report) and White (70.7% White, 6.6% Latino/Hispanic, 9.0% Black, 6.4% Asian or Pacific Island, and 0.5% multi-ethnic) with a wide age range (range 18–78, median = 26 years).

Procedure. After consenting, participants were randomly assigned to one of three priming conditions, and exposed to either an anxiety prime, a security prime, or an avoidance prime. Participants read the following instructions to recall a corresponding close relationship in each condition:

Anxiety Prime. Try to remember a close relationship in which you felt that the other person refused to create a connection as close as you wanted, in which you often feared that the person did not really love you or did not wish to stay with you, in which you felt that you want to merge completely with the person, and this desire sometimes distanced the other person from you.

Security prime. Try to remember a close relationship in which you felt that the goal of getting close to the other person was achieved with relative ease, a relationship in which you felt comfortable being dependent on the person or comfortable with the person being dependent upon you, a relationship in which you did not worry that you would be abandoned or that the person would get too close to you.

Avoidant Prime. Try to remember a close relationship in which you did not feel comfortable getting close to the other person, you had difficulty trusting that person completely, and had difficulty being dependent on the person, a relationship in which you felt tense when the person got too close and often felt as though the person wanted a relationship more intimate than what you were ready for.

After the prime, participants were instructed to describe the relationship in detail for about 3 min. Then they completed the same measures we used in Study 1 to measure trust in AI, experience with AI, and familiarity with AI, followed by some demographic questions. Participants were fully debriefed after the experiment and thanked for their participation. Again, an exploratory factor analysis of the four items measuring trust in AI showed a strong single factor that explained 85.84% of the variances. Therefore, we computed a summed scale score ($\alpha = 0.96$).

2.2. Results

Table 3 lists the descriptive data for trust in AI by prime type. To test whether prime affected people's trust in AI, we ran a hierarchical regression with trust in AI as the dependent variable and prime and age as predictors (see Table 4 for the unstandardized regression coefficients for all predictors). In the first step of the regression, we entered age and familiarity with AI. The distribution of age was again positively skewed, with most of the people falling between 18 and 60, and a few outliers above 60 (see Fig. 2). Similar to Study 1, to resolve this issue we used a log transformation of age in the analysis. In the second step, we entered the prime type. Prime type (security, anxiety, avoidance) was entered as two dummy variables, one (Condition 1) contrasting Attachment Anxiety (−1) with Security (1), and the other (Condition 2) contrasting

Table 3
Descriptive data for trust in AI by condition–Study 2.

Condition	Trust in AI		
	N	M	SD
Anxiety	123	2.91	1.32
Security	122	3.30	1.33
Avoidance	129	3.09	1.37

Table 4
Regression coefficients for trust in AI as a function of age, condition 1, and condition 2—Study 2.

Predicting variables	Step 1	Step 2
Intercept	4.291	4.328
Age	−1.459**	−1.462**
Familiarity	.335**	.324**
Condition 1		.191*
Condition 2		-.049
R^2	.093	.104
R^2 change	.093	.011

Note: Dummy variable Condition 1 contrast Security with Anxiety (1 = Security, −1 = Anxiety); Condition 2 contrast Security and Avoidance condition (1 = Security and −1 = Avoidance). All coefficients represent unstandardized regression coefficients. *p < .05. **p < .001. n = 357.

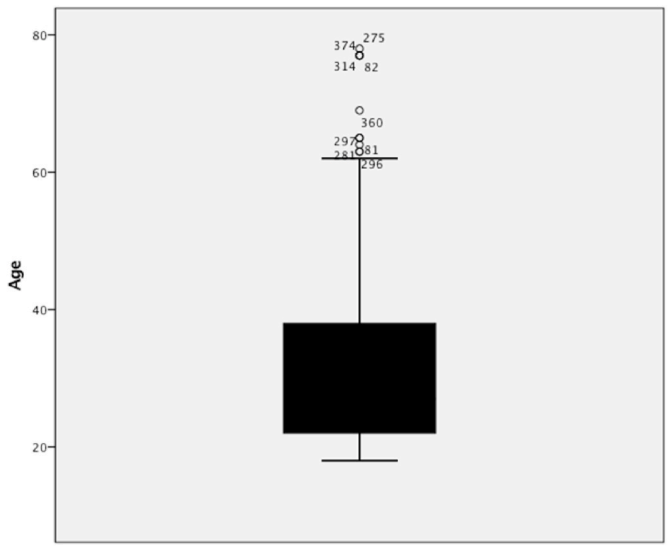

Fig. 2. The boxplot for age –Study 2.

Attachment Avoidance (−1) with Security (1).

The regression analysis revealed a main effect for age, such that the older people were the less they trusted AI. However, it should be noted that the age of our sample was not representative of the full range from 18 to 78, instead, it represented mainly people between the ages of 18

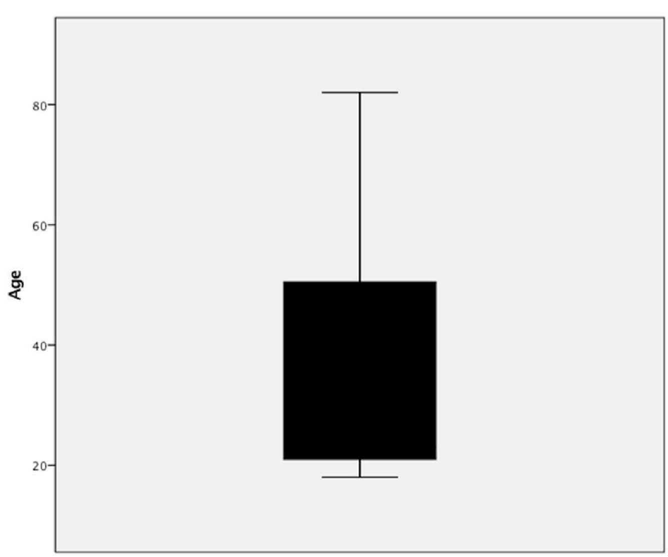

Fig. 3. The boxplot for age –Study 3.

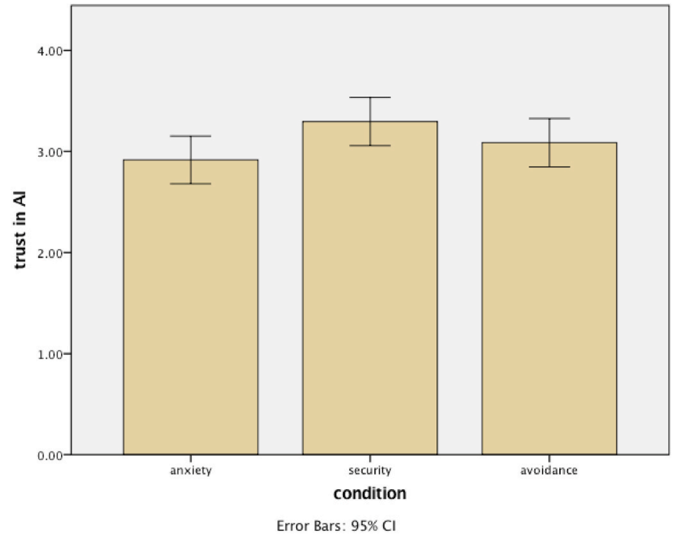

Fig. 4. Trust as a function of prime—study 2.

and 60. Similar to Study 1, we found a significant main effect of familiarity with AI, such that as the familiarity with AI increased, participants were more likely to trust AI. In line with our prediction and with the results of Study 1, the analysis also revealed a main effect for Condition 1, such that people in the Anxiety condition reported lower trust in AI than people in the Security condition (see Fig. 4). This effect was significant even when we included age and familiarity with AI as covariates.

3. Study 3

Study 2 demonstrated that priming attachment anxiety decreased trust in AI. Together Studies 1 and 2 provide convergent validity to the proposition that attachment anxiety is negatively correlated with trust in AI. Study 2 has a few limitations though. First and foremost, looking at Fig. 4, it seems that not only does anxiety priming decrease people's trust in AI, but attachment security increases trust in AI. Study 2, however, does not allow us to directly test whether security indeed increases trust (as we do not have a neutral control condition). Figuring out whether security priming can increase trust is important because if this is the case, it means security could potentially serve as an intervention to increase trust in AI more generally and outside the laboratory. To overcome the limitation of Study 2, in Study 3, we again used an attachment security priming, however, we compared security priming to a neutral priming condition.

A secondary goal of Study 3 was to show that the effects of security could not be ruled out as simply an increase of positive affect. Past work has demonstrated a significant positive association between attachment security and positive affect (see Hazan & Shaver, 1994). At the same time, scholars showed that the effects of security priming could not be attributed to a mere increase in positive affect (Mikulincer et al., 2001). Therefore, while attachment security has a positive mood component to it, it activates more than just positive mood (Canterberry & Gillath, 2013). To examine whether the effects of security priming on trust in AI can be attributed to positive affect, we introduced another control condition in Study 3–positive-affect enhancement condition. We predicted that priming attachment security will lead to higher trust in AI compared to neutral priming, whereas priming positive affect will not.

Finally, in Study 2 we did not include attachment style, to examine whether the effects of priming will take place above and beyond people's attachment style. To deal with this limitation we added an attachment trait measure to Study 3. We predicted that security priming will increase trust regardless of one's attachment style in line with past research (Gillath & Karantzas, 2019).

3.1. Method

Participants. We used a between-subjects design with three conditions. Based on the same power analysis as in Study 2 (Faul et al., 2007) we recruited a total of 272 participants. Participants were recruited in the same way as in Study 1. Participants were mostly females (178 females, 94 males) and White (78.7% White, 5.1% Latino/Hispanic, 2.9% Black, 8.3% Asian or Pacific Island, and 3.2% multi-ethnic) with a wide age range (range 18–82, median = 23 years).

Procedure. After consenting, participants were randomly assigned to one of three conditions (Attachment Security, Positive affect, Neutral). In the Security condition, we primed attachment security by requiring participants to recall a security-providing close relationship:

Try to remember a close relationship in which you felt that the goal of getting close to the other person was achieved with relative ease, a relationship in which you felt comfortable being dependent on the person or comfortable with the person being dependent upon you, a relationship in which you did not worry that you would be abandoned or that the person would get too close to you.

In the positive affect condition, we asked participants to imagine winning an all-expenses-paid vacation to the Bahamas. In the neutral condition, we asked participants to remember a time when they did some mundane task with an acquaintance of theirs, like buying products in the store or studying in the library with a classmate. Participants were instructed to describe the relationship or the event in detail for about 3 min: "*You may refer to external events, behaviors of the parties involved, and internal feelings, thoughts, emotions, desires, and the like.*"

After the prime, participants completed the same measures we used in Studies 1 and 2 to measure trust in AI, attachment style (anxiety: $\alpha = 0.89$; avoidance: $\alpha = 0.80$), experience with AI, familiarity with AI, and finally some demographic questions. Participants were fully debriefed after the experiment. Again, an exploratory factor analysis of the trust in AI measure showed a strong single factor that explained 89.71% of the variances. Therefore, we computed a summed scale score ($\alpha = 0.95$).

3.2. Results and discussion

Table 5 lists the descriptive data for trust in AI by the prime type (see also Fig. 5). To test whether prime, attachment style, and age affected people's trust in AI, we ran a hierarchical regression with trust in AI as the dependent variable and the other variables as predictors (see Table 6 for the unstandardized regression coefficients for all predictors). In the first step of the regression, we entered age and familiarity with AI. Similar to Study 1 and Study 2, we entered the log transformation of age because of the positively skewed distribution of age (see Fig. 3). In the second step, we entered standardized scores of attachment anxiety and attachment avoidance, and prime. Prime (Security, Positive, Neutral) was entered as two dummy variables, one (Condition 1) contrasting Neutral (−1) with Positive (1), and the other (Condition 2) contrasting Neutral (−1) and Security (1). In the next step, we entered all the two-way interactions between attachment anxiety, attachment avoidance, Condition 1, and Condition 2. In the fourth step, we entered the three-way interactions between attachment anxiety, attachment avoidance, and Condition 1, and between attachment anxiety, attachment avoidance, and Condition 2.

Table 5
Descriptive data for trust in AI by condition– Study 3.

Condition	Trust in AI		
	N	M	SD
Positive affect	97	2.83	1.19
Security	91	2.97	1.08
Neutral	89	2.63	1.08

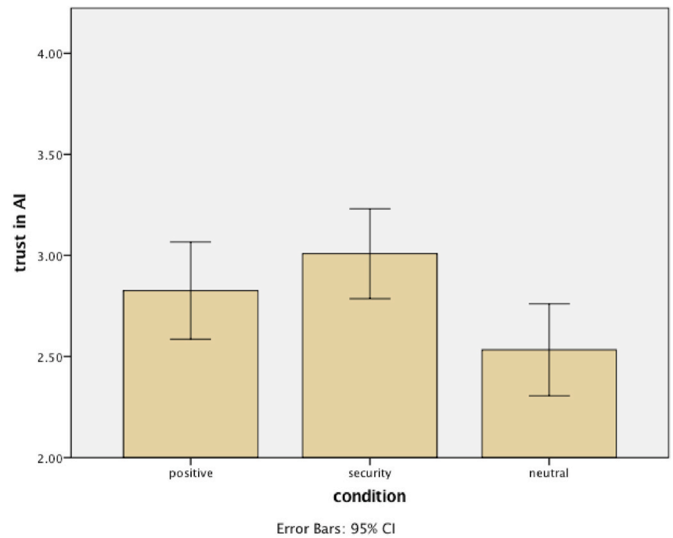

Fig. 5. Trust as a function of prime—study 3.

Table 6
Regression coefficients for trust in AI as a function of age, attachment anxiety, attachment avoidance, condition 1 and condition 2—Study 3.

Predicting variables	Step 1	Step 2	Step 3	Step 4
Intercept	3.255	3.111	3.097	3.085
Age	-.1.020**	-.938*	-.918*	-.927*
Familiarity	.402**	.411**	.407**	.418**
Avoidance		.113	.114	.121
Anxiety		.020	.003	.007
Condition 1		.039	.033	.025
Condition 2		.198*	.196*	.187*
Avoidance*anxiety			-.068	-.061
Condition1*avoidance			-.125	-.127
Condition1*anxiety			-.063	-.057
Condition2*avoidance			-.016	-.015
Condition2*anxiety			.137	.155
Condition1*anxiety*avoidance				.023
Condition2*anxiety*avoidance				.082
R^2	.157	.190	.207	.212
R^2 change	.157	.033	.017	.005

Note: Dummy variable Condition 1 contrast Positive affect and Neutral condition (1 = Positive affect and −1 = Neutral); Condition 2 contrast Security and Neutral condition (1 = Security and −1 = Neutral). All coefficients represent unstandardized regression coefficients. *$p < .05$. **$p < .001$. n = 257.

The regression analysis revealed a main effect for age, such that older people were less likely to trust AI. We also found a significant main effect of familiarity with AI, such that as the familiarity with AI increased, participants were more likely to trust AI. The analysis also revealed a main effect for Condition 2, such that people in the Security condition reported higher trust in AI than people in the Neutral condition. Condition 1 did not predict trust in AI, people in the Positive affect condition did not report higher trust in AI than people in the Neutral condition. The results suggest that only security prime enhancement increases trust in AI, and this effect could not be attributed to positive affect. The effects of security enhancement remained significant even when controlling for age, familiarity with AI, and dispositional attachment anxiety and avoidance. No other main effects or interactions were significant.

4. General discussion

In three studies, we examined the links between attachment style and trust in AI. In Study 1, we found that attachment anxiety but not avoidance predicted less trust in AI. These results held even when we controlled for the potential role of neuroticism and self-esteem. In Study

2, we found that enhancing the sense of attachment anxiety resulted in a decrease in trust in AI. Study 2 also seemed to suggest that enhancing attachment security could increase trust in AI. In Study 3, we indeed found that enhancing the sense of attachment security via priming as compared with exposure to a neutral prime, resulted in increased trust in AI. Overall, our findings highlight the need to further study the affective route to boost trust, the importance of personality traits in predicting this trust, and the potential of attachment security enhancement to serve as an intervention to increase trust in AI.

Study 1 is the first study, to our knowledge, to demonstrate that attachment style can be used to predict how people feel about AI. More specifically, we showed that similarly to the associations found between attachment style and human-human relations or human-robot relations, attachment style was found to predict people's trust in non-human AIs. These findings are highly important as although most people might not encounter machines and robots in their daily lives, they have already or will likely encounter some sort of AI in the near future (like personal helpers; Siri, Google Home). Finding models and theories that are relevant for human-AI relations can help improve these relations and the overall functioning and performance of human-AI teams (Chakraborti et al., 2017). Furthermore, these findings support the suggested link between affective factors, such as attachment, and trust in AI.

Importantly, we showed that our findings were unique to attachment style. That is, our results held even when we added self-esteem and neuroticism to the model. Although previous research on the Big Five (e.g., Noftle & Shaver, 2006), and Self-esteem (Meyers, 1998) showed correlations between these personality traits and attachment style, we demonstrated in Study 1, that even when one controls for these traits, attachment style still predicts trust in AI. Study 1 also revealed effects for age and familiarity with AI, such that older adults and people less familiar with AI were less likely to trust it. The age effects were similar throughout our studies. Although the actual age range in some of the studies was a bit limited and the distribution was skewed to the right, once we corrected the skewness the results were similar. These findings suggest that being younger and more familiar with AI are positively correlated with trusting AI. Our findings also suggest that security priming might especially help older adults who are less familiar with AI.

Although we expected attachment insecurity overall to predict lower trust, attachment avoidance was not correlated with trust in AI. This was surprising, as avoidance is correlated with less trust among humans (Pistole, 1993). Perhaps relations with AIs are overall less threatening than relationships with humans for avoidant people, and therefore, avoidance does not play as big of a role in predicting trust in AI. This proposition is supported by studies showing that avoidant people avoid close others, but have no issues getting closer to strangers (Fraley & Shaver, 1998). Perhaps AIs are likely to be perceived more like strangers than as close others by avoidant people. This is a potential topic for future research.

Study 2 provided further support to our proposition that attachment insecurity and specifically attachment anxiety is associated with less trust in AI, and that enhancing the sense of attachment anxiety results in decreased trust in AI. We used an experimental design to compare attachment anxiety priming to attachment security and attachment avoidance priming. As predicted, we found that anxiety priming resulted in lower trust in AI as compared with security. These findings are similar to findings linking anxiety and lack of trust in human relations (Pistole, 1993). This lack of trust might be due to the tendency of anxiously attached people to experience insecurity about others' responsiveness and availability. The comparison between security and avoidance priming did not yield a significant effect.

On the one hand, these findings support our overall proposition that attachment predicts and impacts trust in AI. On the other hand, these results could not provide us with a clear picture regarding the effectiveness and strength of security priming. To further examine the effect of security priming we conducted a third study that allowed us to compare security priming to a neutral control condition.

In Study 3, we used a similar design as in Study 2, however this time we added a neutral control condition and a positive-affect control condition. Supporting our predictions, and the preliminary findings of Study 2, Study 3 showed that enhancing the sense of attachment security via priming leads people to trust AI more, as compared to neutral priming. Increasing the sense of security makes people feel more safe and secure, which in turn allows them to lower their defenses and trust others (including AIs) more. We also showed that the effects of security priming on trust in AI could not be explained away as simply an increase in positive affect. These findings support the idea that attachment security could be used as an intervention to increase trust in AI and that primes focusing only on positive affect will not achieve the same outcome. Our findings are in line with the work of Mikulincer et al. (2001) showing that security priming goes above and beyond positive affect to impact people's relationships and behavior more broadly.

Limitations and future directions. There are three main limitations to our studies. First, our results are based on guided imagery (imagining different scenarios), which lowers their validity. Self-reports suffer from various biases such as social desirability and common source, so future studies should include other, non-self-report measures. That said, the examples we included in our scenarios were all based on plausible real-world interactions and similar to scenarios used in other studies (Dujmovic, 2017; Elangovan et al., 2007; Xie & Peng, 2009), some of which our participants already encountered and are highly familiar with. Furthermore, our findings are in line with the findings of Birnbaum et al. (2016) showing that humans ascribe social intentions to non-human robots and can use them as a source of consolation and security. Here we added that humans can also trust (or not) AIs, and this trust (or lack of) can be predicted and manipulated using attachment style. Future studies should include more scenarios and increase the representativeness of the trust measure, and potentially assess trust as two separate domains (affective and cognitive) rather than one amalgam as we did here.

A second limitation has to do with the control conditions we used. Although the three studies and especially the experimental studies covered a wide range of controls, it would be helpful for future research to include both a neutral condition and insecurity primes when examining the effects of attachment security on trust in AI. That will allow researchers to directly compare anxiety and avoidance to each other, rather than each of them to a secure condition, and further understand the impact of insecurity on trust. Additional control could include self-esteem priming.

A third and related limitation has to do with the lack of a neutral condition in Study 2. Although supporting our prediction (anxiety decreases trust and/or security increases it), the results cannot provide conclusive evidence; in other words, we do not know if anxiety decreased trust, security increased trust, or both. Looking at Figs. 4 and 5, it is clear that security priming results in higher trust in both studies, however, without a neutral control condition similar to the one we had in Study 3 we cannot reach a conclusion on which of these three options was supported in Study 2. Future studies should include a neutral prime with the insecurity controls.

There are a few directions that one could develop from the current findings. The first has to do with where our work intersects with that of XAI. Questions such as: What constitutes a trusted explanation? What types of explanations are most likely to enhance trust? As well as, How can explanations be tailored by attachment style? Follow from our work and will benefit from further examination. A different direction has to do with the moral implications of our procedure and findings—questions such as why do people have to trust AI? And what are the implications for consumers, when they are primed to trust AI but do not want to? A third direction is related to the previous one and has to do with the effects of priming security outside of one's awareness (subliminal priming). What would be the effects of such priming? And what would be the moral implications of doing that? These questions and others would benefit from future research on the intersection of attachment, affective trust, and AI. A final direction has to do with the manipulation we used in the current set of studies. In light of our findings, we suggest that both researchers and industry leaders consider using similar manipulations as we had here or other ways to decrease anxiety and increase security as means to help increase trust in AI, especially in the initial stages of adoption. This could be achieved by for example, making AI more similar to one's loved ones, or by making people think about their close relationships.

Conclusion. Overall, the results of the three studies reported above highlight the usefulness of personality traits for predicting trust in AI, the ability to generalize from trust models among humans to AI, and the potential of security enhancement to serve as an intervention to increase trust in AI. Our studies suggest that the way humans treat each other is potentially similar to the way they interact with or think about AIs. Our paper is one of the only papers to link attachment style and trust in AI, and the only one to show that security priming can boost trust. Importantly, we show that these effects on trust are not merely human-focused, but also apply to non-humans. Finally, we showed that general trust in AI, which in the past was boosted mainly via the cognitive route, can also be increased using affective manipulations. Our results directly affect the adoption and success of AI technologies in individuals and in workplace and public settings. They also inform the design of trusted AI systems. Finally, they form a new complementary focus to the traditional focus on cognitive trust in AI and explainable AI.

CRediT author statement

Omri Gillath: Conceptualization, Methodology, Writing- Original draft preparation, Revising paper, Supervision **Ting Ai:** Formal analysis, Methodology, Writing results, revising **Michael Branicky:** Conceptualization, Writing-Reviewing and Editing, **Shawn Keshmiri:** Writing-Reviewing and Editing, **Robert B. Davison:** Conceptualization, Writing-Reviewing and Editing, and **Ryan Spaulding:** Writing-Reviewing and Editing.

References

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. N. (1978). *Patterns of attachment: A psychological study of the strange situation*. Psychology Press.

Akash, K., Hu, W., Reid, T., & Jain, N. (2017). Dynamic modeling of trust in human-machine interactions. In *2017 American control conference*. ACC).

Andriole, S. (2018). AI: The good, the disruptive, and the scary. *Cutter Business Technology Journal, 31*, 6–11.

Bartz, J. A., & Lydon, J. E. (2004). Close relationships and the working self-concept: Implicit and explicit effects of priming attachment on agency and communion. *Personality and Social Psychology Bulletin, 30*, 1389–1401.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*, 497–529.

Berg, A., Buffie, E. F., & Zanna, L. F. (2018). Should we fear the robot revolution? (The correct answer is yes). *Journal of Monetary Economics, 97*, 117–148.

Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., & Sass, O. (2016). What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior, 63*, 416–423.

Bliss, J. P., Washington, M. D., & Fuller, B. S. (1994). Reversal of the cry-wolf effect: An investigation of two methods to increase alarm response rates. *Proceedings of the Human Factors and Ergonomics Society - Annual Meeting, 38*, 968. https://doi.org/10.1177/154193129403801547, 968.

Bloomberg, J. (2018). *Why people don't trust artificial intelligence: It's an 'explainability' problem*. Forbes. https://geneticliteracyproject.org/2018/09/26/why-people-dont-trust-artificial-intelligence-its-an-explainability-problem/.

Bowlby, J. (1982), 1969. *Attachment and loss* (original ed, Vol. 1). London: Random House.

Brachman, R. J. (2006). AI more than the sum of its parts. *AI Magazine, 27*, 19–19.

Braynov, S., & Sandholm, T. (2002). Contracting with uncertain level of trust. *Computational Intelligence, 18*, 501–514.

Canterberry, M., & Gillath, O. (2013). Neural evidence for a multifaceted model of attachment security. *International Journal of Psychophysiology, 88*, 232–240.

Cassidy, J., & Shaver, P. R. (2018). *Handbook of attachment theory, research, and clinical applications*. New York, London: The Guilford Press.

Center for the Governance of AI. Future of humanity institute, university of Oxford. https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/index.html.

Chakraborti, T., Kambhampati, S., Scheutz, M., & Zhang, Y. (2017). *Ai challenges in human-robot cognitive teaming*. arXiv preprint arXiv:1707.04775.

Chang, L. (2016). New study suggests Americans don't trust AI systems. Digital Trends. https://www.digitaltrends.com/cool-tech/ai-system-trust/.

Cook, J., & Wall, T. (1980). New work attitude measures of trust, organizational commitment and personal need non-fulfilment. *Journal of Occupational Psychology, 53*, 39–52.

Crandall, J. W., Oudah, M., Tennom, Ishowo-Oloko, F., Abdallah, S., Bonnefon, J. F., , … Rahwan, I., et al. (2018). Cooperating with machines. *Nature Communications, 9*, 233. https://doi.org/10.1038/s41467-017-02597-8

Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In *Robot law*. Edward Elgar Publishing.

Dikmen, M., & Burns, C. M. (2017). Trust in autonomous vehicles: The case of tesla autopilot and summon. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*.

DriversEd.com. (2019). DriversEd.com 2019 state of self-driving cars report. https://driversed.com/trending/2019-state-self-driving-cars-report.

Doran, P., Tamma, V., & Iannone, L. (2007). Ontology module extraction for ontology reuse: an ontology engineering perspective. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 61–70).

Dujmovic, J. (2017). Opinion: What's holding back artificial intelligence? Americans don't trust it. at 5:08 a.m. ET https://www.marketwatch.com/story/whats-holding-back-artificial-intelligence-americans-dont-trust-it-2017-03-30. (Accessed 30 March 2017).

Dzindolet, M. T., Pierce, L. G., Beck, H. P., & Dawe, L. A. (2002). The perceived utility of human and automated aids in a visual detection task. *Human Factors, 44*, 79–94.

Elangovan, A. R., Auer-Rizzi, W., & Szabo, E. (2007). Why don't I trust you now? An attributional approach to erosion of trust. *Journal of Managerial Psychology, 22*, 4–24.

Evers, V., Maldonado, H. C., Brodecki, T. L., & Hinds, P. J. (2008). Relational vs. group self-construal: Untangling the role of national culture in HRI. *2008 3rd ACM*. In *IEEE international conference on human-robot interaction (HRI)* (pp. 255–262).

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Fraley, R. C., & Shaver, P. R. (1998). Airport separations: A naturalistic study of adult attachment dynamics in separating couples. *Journal of Personality and Social Psychology, 75*, 1198–1212.

Gillath, O., & Karantzas, G. (2019). Attachment security priming: A systematic review. *Current opinion in psychology, 25*, 86–95.

Gillath, O., Karantzas, G. C., & Fraley, R. C. (2016). *Adult attachment: A concise introduction to theory and research*. Academic Press.

Gillath, O., Sesko, A. K., Shaver, P. R., & Chun, D. S. (2010). Attachment, authenticity, and honesty: Dispositional and experimentally induced security can reduce self-and other-deception. *Journal of Personality and Social Psychology, 98*, 841–855.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B., Jr. (2003). A very brief measure of the big five personality domains. *Journal of Research in Personality, 37*, 504–528.

Groom, V., & Nass, C. (2007). Can robots be teammates? Benchmarks in human-robot teams. *Interaction Studies, 8*, 483–500.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors, 53*, 517–527.

Hart, J., Shaver, P. R., & Goldenberg, J. L. (2005). Attachment, self-esteem, worldviews, and terror management: Evidence for a tripartite security system. *Journal of Personality and Social Psychology, 88*, 999–1013.

Hazan, C., & Shaver, P. R. (1994). Attachment as an organizational framework for research on close relationships. *Psychological Inquiry, 5*, 1–22.

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

Hiolle, A., Cañamero, L., Davila-Ross, M., & Bard, K. A. (2012). Eliciting caregiving behavior in dyadic human-robot attachment-like interactions. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*, 3.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*, 407–434.

Hughes, J. S., Rice, S., Trafimow, D., & Clayton, K. (2009). The automated cockpit: A comparison of attitudes towards human and automated pilots. *Transportation Research Part F: Traffic Psychology and Behaviour, 12*, 428–439.

Jeffries, F. L., & Reed, R. (2000). Trust and adaptation in relational contracting. *Academy of Management Review, 25*, 873–882.

Jiang, X., Khasawneh, M. T., Master, R., Bowling, S. R., Gramopadhye, A. K., Melloy, B. J., & Grimes, L. (2004). Measurement of human trust in a hybrid inspection system based on signal detection theory measures. *International Journal of Industrial Ergonomics, 34*, 407–419.

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*, 1306–1317.

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research, 58*, 500–507.

Knowles, B., & Hanson, V. L. (2018). The wisdom of older technology (non)users. *Communications of the ACM, 61*, 72. https://doi.org/10.1145/3179995

Konok, V., Korcsok, B., Miklósi, A., & Gácsi, M. (2018). Should we love robots? The most liked qualities of companion dogs and how they can be implemented in social robots. *Computers in Human Behavior, 80*, 132–142.

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

Liang, Y., & Lee, S. A. (2017). Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics, 9*, 379–384.

Looije, R., Neerincx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies, 68*, 386–397.

Lo, C., Walsh, A., Mikulincer, M., Gagliese, L., Zimmermann, C., & Rodin, G. (2009). Measuring attachment security in patients with advanced cancer: Psychometric properties of a modified and brief experiences in close relationships scale. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer, 18*, 490–499. https://doi.org/10.1002/pon.1417

Lu, S. C., Kong, D. T., Ferrin, D. L., & Dirks, K. T. (2017). What are the determinants of interpersonal trust in dyadic negotiations? Meta-analytic evidence and implications for future research. *Journal of Trust Research, 7*, 22–50.

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*, 277–301.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: A trust building model. *The Journal of Strategic Information Systems, 11*, 297–323.

Meyers, S. A. (1998). Personality correlates of adult attachment style. *The Journal of Social Psychology, 138*, 407–409.

Mikulincer, M. (1998). Attachment working models and the sense of trust: An exploration of interaction goals and affect regulation. *Journal of Personality and Social Psychology, 74*, 1209–1224.

Mikulincer, M., Hirschberger, G., Nachmias, O., & Gillath, O. (2001). The affective component of the secure base schema: Affective priming with representations of attachment security. *Journal of Personality and Social Psychology, 81*, 305.

Moorman, C., Zaltman, G., & Deshpande, R. (1992). Relationships between providers and users of market research: The dynamics of trust within and between organizations. *Journal of Marketing Research, 29*, 314–328.

Morsünbül, Ü. (2018). Attachment and sex with robots: An assessment from mental health perspective. *Current Approaches in Psychiatry, 10*, 417–429. https://doi.org/10.18863/pgy.363669

Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior, 72*, 432–440.

Noftle, E. E., & Shaver, P. R. (2006). Attachment dimensions and the big five personality traits: Associations and comparative ability to predict relationship quality. *Journal of Research in Personality, 40*, 179–208.

Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: A survey. *Artificial intelligence for human computing*. Berlin, Heidelberg: Springer.

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

Pegasystems. (2018). What consumers really think about AI: A global study. https://www.ciosummits.com/what-consumers-really-think-about-ai.pdf.

Pistole, M. C. (1993). Attachment relationships: Self-disclosure and trust. *Journal of Mental Health Counseling, 15*, 94–106.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49*, 95–112.

Robinson, H., MacDonald, B., & Broadbent, E. (2014). The role of healthcare robots for older people at home: A review. *International Journal of Social Robotics, 6*, 575–591.

Rosenberg, M. (1965). Rosenberg self-esteem scale (RSE). *Acceptance and commitment therapy*. *Measures package, 61*(52), 18.

Rossi, F. (2018). Building trust in artificial intelligence. *Journal of International Affairs, 72*, 127–134.

Russell, S., & Norvig, P. (2020). *Artificial intelligence: a modern approach* (4th ed). Google Inc.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67.

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller. (2019). In *Explainable AI: interpreting, explaining and visualizing deep learning, 11700*. Springer Nature.

Sanders, T., Oleson, K. E., Billings, D. R., Chen, J. Y., & Hancock, P. A. (2011). A model of human-robot trust: Theoretical model development. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 55, pp. 1432–1436). Sage CA: Los Angeles, CA: SAGE Publications.

Sarkar, S., Araiza-Illan, D., & Eder, K. (2017). Effects of faults, experience, and personality on trust in a robot Co-worker. Available Online at: https://arxiv.org/abs/1703.02335.

Siau, K., & Wang, W. (2018). Trusting artificial intelligence in healthcare. In *24th americas conference on information systems* (New Orleans).

Simmons, B. L., Gooty, J., Nelson, D. L., & Little, L. M. (2009). Secure attachment: Implications for hope, trust, burnout, and performance. *Journal of Organizational Behavior, 30*, 233–247.

Stormont, D. P. (2008). Analyzing human trust of autonomous systems in hazardous environments. In *Proc. Of the human implications of human-robot interaction workshop at AAAI* (pp. 27–32).

Sugiyama, S., & Vincent, J. (2013). Social robots and emotion: Transcending the boundary between humans and ICTs. *Inter, 1*, 1–6.

SurveyUSA. (2020). AV Perceptions and Attitudes. Partners for automated vehicle education. https://pavecampaign.org/news/pave-poll-americans-wary-of-avs-but-say-education-and-experience-with-technology-can-build-trust/.

Thau, S., Crossley, C., Bennett, R. J., & Sczesny, S. (2007). The relationship between trust, attachment, and antisocial work behaviors. *Human Relations, 60*, 1155–1179.

Towers-Clark, C. (2019). 80% of people don't trust AI with money - how can we fix its image? *Forbes*. https://www.forbes.com/sites/charlestowersclark/2019/01/15/80-of-people-dont-trust-ai-with-money-how-can-we-fix-its-image/#2d7994b0759a.

Tussyadiah, I. P., Zach, F. J., & Wang, J. (2017). Attitudes toward autonomous on demand mobility system: The case of self-driving taxi. *Information and Communication Technologies in Tourism, 2017*, 755–766.

Verberne, F. M. F., Ham, J., & Midden, C. J. H. (2012). Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human Factors, 54*, 799–810.

Wang, P. (2008). *What Do You Mean by "AI"?. Artificial General Intelligence* (pp. 362–373).

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

Webber, S. S. (2008). Development of cognitive and affective trust in teams: A longitudinal study. *Small Group Research, 39*, 746–769.

Xie, Y., & Peng, S. (2009). How to repair customer trust after negative publicity: The roles of competence, integrity, benevolence, and forgiveness. *Psychology and Marketing, 26*, 572–589.

Yokoi, R., & Nakayachi, K. (2018). The effects of shared policy of medical treatment on trust in artificial intelligence. *Japanese Journal of Social Psychology, 34*, 16–25.

You, S., & Robert, L. P. (2018). Emotional attachment, performance, and viability in teams collaborating with embodied physical action (EPA) robots. *Journal of the Association for Information Systems, 19*, 377–407.

You, S., & Robert, L. P. (2019). Trusting robots in teams: Examining the impacts of trusting robots on team performance and satisfaction. In *Proceedings of the 52nd Hawaii international conference on system sciences (HICSS 2019), jan 8-11*. USA: Maui, HI.

Zhang, B., & Dafoe, A. (2019). *Artificial intelligence: American attitudes and trends*.