"Do you still trust me?" Effects of Personality on Changes in Trust during an Experimental Task with a Human or Robot Partner

Anthony M. Gibson
Consortium of Universities
anthony.gibson.9.ctr@us.af.mil

Gene M. Alarcon
Air Force Research Laboratory
gene.alarcon.1@us.af.mil

Sarah A. Jessup
Air Force Research Laboratory
sarah.jessup.ctr@us.af.mil

August Capiola
Air Force Research Laboratory
august.capiola.1@us.af.mil

Abstract

In the current study, we investigated the effects of dispositional variables on self-reported trust and suspicion perceptions of one's partner in a maze-running task. Dispositional variables interact with situational variables in expressing behaviors. In order to test the effects of three dispositional variables (i.e., dispositional trust, dispositional distrust, and dispositional suspicion) on self-reported trust and suspicion perceptions towards a partner (a human or a Nao robot), we ran two discontinuous growth models. Overall, we found that participants' trust towards their partner decreased when the partner engaged in untrustworthy behaviors as expected. In addition, changes in trust perceptions towards the partner were predicted by participants' level of dispositional trust. These results have implications for studying the effects of dispositional variables on context-dependent trust perceptions within the trust process.

1. Introduction

Interpersonal trust affects the degree to which people develop and maintain various healthy social relationships, including both romantic [1] and work [2] relationships. Interpersonal trust researchers began studying trust in human-human interactions. In this context, Mayer, Davis, and Schoorman [3] defined trust as a person's willingness to be vulnerable to another, without the trustor monitoring that trustee's actions. Human-human trust has been studied extensively in the workplace, and researchers have stated that trust has shown increased importance in the workplace with the emergence of work teams [4] and increased diversity in the workforce [3].

Indeed, Colquitt, Scott, and LePine [5] used meta-analytic techniques and found that dispositional trust, a general willingness to trust others, had both direct effects on job performance (i.e., task performance, citizenship behaviors, and counterproductive behaviors) and indirect effects on performance via perceptions of trustworthiness and trust intentions. In their seminal article, Mayer et al. [3] delineated between three factors that compose the trust process (i.e., dispositional trust, perceived trustworthiness, and trust intentions).

Recently, researchers have focused on the role of dispositional trust in the trust process when humans interact with automation [6, 7, 8]. In their extensive review, Hoff and Bashir described trust applied to interactions with technology as "*trust in automation*." Generally, researchers have found that human users are more likely to use, and less likely to monitor, automation they trust, especially under uncertain conditions [6, 9, 10].

We should note, however, that trust is dynamic and can change based on the nature of human-automation interactions [6]. Hoff and Bashir described three separate layers of trust variability, which include: (a) dispositional trust, (b) situational trust, and (c) learned trust. Dispositional trust was defined as one's tendency to be willing to be vulnerable to automation, whereas situational trust was defined as one's willingness to be vulnerable to automation in specific contexts. Learned trust refers to the attitudes the trustor has towards the automation, which is gained from prior experience or present interactions with the system. If the person has no prior experience with the system, then he or she must rely on the system's current performance to calibrate dynamic learned trust [6]. In fact, prior research has found that both consistency (i.e., reliability) and predictability of automation influence trust [11, 12, 13, 14].

H͓CSS

Although characteristics of the automation (e.g., reliability) affect the extent to which humans trust it, dispositional trust also affects the degree to which humans rely on automation [6, 14]. Specifically, Hoff and Bashir noted that dispositional trust has the greatest influence on learned trust before the person interacts with the automation. Similarly, researchers within the human-human interpersonal trust domain have found that dispositional trust has the greatest effect on the trust process during initial interactions (i.e., when the trustor has limited interactions with the trustee; [15, 16, 17]). Within the human-automation literature, researchers have empirically demonstrated that dispositional trust influences perceived trustworthiness prior to interaction with automation, but not after extended exposure [14]. Thus, participants scoring higher in dispositional trust might be more likely to report more trustworthiness towards the referent early in the experiment.

Stemming from Rotter's [18] early work, Mayer and Davis [19] published the Propensity to Trust scale, a measure of dispositional trust, which has been used widely across the organizational [5] and the personality [15, 20] literatures. In addition to the Propensity to Trust scale, an IPIP NEO-PI-R Trust facet of Agreeableness [21] measures the extent to which humans have a predisposition to trust others generally. The items that compose the Trust facet scale are more general than those in the Propensity to Trust scale [19]. For example, a sample item from the IPIP NEO-PI-R Trust facet scale reads, "Trust others," whereas a sample item from Mayer and Davis's scale reads, "Most sales people are honest in describing their products." Because the IPIP NEO-PI-R Trust facet scale comprises more general items than the Propensity to Trust scale, Mayer and Davis's scale may predict specific trust intentions in the trust process better than IPIP NEO-PI-R Trust. In other words, because items from the IPIP NEO-PI-R Trust facet scale do not include specific referents, the scale may be more distal to trust intentions than items from Mayer and Davis' scale. In comparison, Mayer and Davis' scale may predict criterion more proximal to the context-dependent trust intentions. Similar research from the human factors literature shows that more contextually relevant and proximal items measuring propensity to trust automation had greater predictive validity of trust outcomes of interest than scales comprising more general items [22]. Constructs other than dispositional trust, however, may be important predictors of trust-relevant outcomes, too.

Dispositional distrust corresponds to a person's predisposition to assume others are untrustworthy and have intentions to harm the trustor [23].

Although some researchers have made clear that trust and distrust reside on a continuum [24], others have claimed that distrust is a separate construct from trust [25]. Chang and Fang [23] found that participants' context-dependent trust and distrust influenced internet-relevant behaviors. Specifically, the effects of context-dependent distrust on the prevention of risky behavior were stronger compared to the effects of trust on the promotion of low-risk behaviors. As such, dispositional trust and dispositional distrust may account for unique variance in trust-related outcomes. Thus, we tested whether a measure of dispositional distrust predicts trust-related perceptions after controlling for the effects of dispositional trust.

Similar to the trust-distrust debate, prior research has suggested that suspicion is a separate construct from trust and distrust [26, 27]. Suspicion in the moment comprises uncertainty, cognitive activity, and perceived mal-intent. Similarly, Bobko et al. defined dispositional suspicion as the degree to which a person engages in cognitive processing and perceives mal-intent when interacting in ambiguous circumstances in general. To test this claim, Capiola et al. [28] found some support that dispositional trust and suspicion are separate, but related, constructs. Furthermore, dispositional trust and suspicion explained unique variance in outcomes of interest [28]. As such, dispositional suspicion may explain variance in context-dependent trust and suspicion after controlling for dispositional trust and distrust.

1.1. The current study

The current study investigated the effects of dispositional trust, dispositional distrust, and dispositional suspicion on trustworthiness perceptions and context-dependent suspicion during a laboratory task, in which participants were randomly assigned to interact with a human or automated partner. We tested these effects by studying repeated measurements across six time points on two trust-related measures. Stated simply, we used individual difference variables to explain inter-individual and intra-individual differences in trust-related perceptions over time. We sought to answer the following research questions:

- RQ1: Are there differences in perceived trustworthiness and suspicion across partner conditions?
- RQ2: What individual difference variables are the best predictors of changes in partner trustworthiness perceptions before and after a trust violations?

- RQ3: What individual difference variables are the best predictor of changes in suspicion perceptions before and after a trust violations?

2. Method

2.1. Participants

Participants (N = 49) were undergraduates (57% female) from a medium-sized Midwestern university. Participants' average age was 23.04 years (SD = 4.62). Participants completed the study in exchange for monetary compensation.

2.2. Experimental task

Participants completed the Checkmate task [29], a modified version of the investor/dictator game [30]. Researchers have used the investor/dictator game to measure trust amongst a dyad, with the objective being to earn money throughout the session. The dyad consists of an investor (i.e., the person who controls the amount of money given to the dictator) and the dictator (i.e., the person who decides how much money to return to the investor). The game can be played across multiple rounds, and the amount of money shared during each round between the dyad can be interpreted as an act of trust between the partners. The investor/dictator game was modified to remove the transparency of how much money each person has at any point in the game [29]. Because trust is defined as the willingness to be vulnerable to another's actions independent of the ability to monitor the other's behavior, [29] created the Checkmate task to embed ambiguity into the game (i.e., removing the trustor's ability to monitor trustee's behavior directly).

In the current version of Checkmate, the participant was always assigned to the role of banker (i.e., the investor in the investment/dictator game). A robot [or a human confederate] was always assigned the role of runner (i.e., the dictator in the investment/dictator game). The participant loaned money to the runner over six experimental rounds, and the runner ran a virtual maze to collect as many boxes as possible to increase monies. The banker started with $50 USD in a virtual bank account. The banker loaned money to the runner each round in anticipation of earning interest on his investment.

At the beginning of each round, the runner promised to return the initial loan and 50% of the earnings to the banker. The banker was notified via a pop-up message of the risk level and the amount of money the runner promised to return. Then, the banker selected an amount to loan to the runner: Small ($1 - $7), Medium ($4 - $10), or Large ($7 - $13). Money was then transferred into the runner's virtual wallet. Once the maze-running task started, the banker watched a top-down video of the runner's progress. The runner was allotted two minutes to collect as many boxes as possible. After the maze-running task completed, the runner decided the amount of money to return to the banker. The banker received a pop-up message of the exact amount of money the runner decided to return.

The steps outlined above were repeated over six rounds. Participants were informed that all money exchanged in the task represented real money. The amount of money the banker had in his/her virtual bank at the end of the session belonged to the banker, and the earnings were paid out in the form of cash, rounded up to the nearest quarter. The participant also received a $30 USD gift card for participating.

2.3. Manipulations

Amongst the dyad, the participant was always the banker, and the runner was either a human confederate or the Nao robot. The Nao robot is a commercially-available humanoid robot with anthropomorphic features (e.g., eye gaze, natural movements). Users can program the robot to say various statements, as well as address people by name. In the current study, the robot introduced himself and stated he was ready to perform the task.

We sought to examine differences in trust and suspicion across time depending on whether participants interacted with a human or the Nao robot. If participants were assigned to the human condition, the participants and the confederate introduced themselves, and the experimenter stated that the two would complete the study simultaneously. If participants were assigned to the robot condition, we described that the robot behaves in a self-preserving manner, which would be beneficial if the robot operated in a dangerous environment in contexts outside the experiment. We included this backstory to provide a reasonable explanation as to potential reasons the robot may engage in distrust behaviors. All participants' experiences with the runner were identical. Within the first three time points, the runner returned the amount of money that was promised (i.e., a trustworthy behavior). At the fourth and fifth time points, the runner returned less money than was promised to the banker (i.e., a distrust behavior). In

the final time point, the runner engaged in another trusting behavior by returning the amount promised.

All the runner's data, including maze performance, returning of investment to banker, was prerecorded. This level of control allowed a focus on the way that participants trusted their partner. However, participants were led to believe they were interacting with the robotic or human partner in real-time.

2.4. Measures

2.4.1. Dispositional Trust. Dispositional trust was assessed using two measures: Propensity to Trust scale [19] and the IPIP NEO-PI-R Agreeableness Trust facet [21].

2.4.1.1. *Propensity to Trust scale*. The Propensity to Trust scale [19] is an 8-item scale used to measure one's general tendency to trust others. Two items (i.e., "One should be very cautious of strangers" and "These days, you must be alert or someone is likely to take advantage of you") were removed during data cleaning, due to these items reducing the internal consistency of the scale. A sample item is, "Most experts tell the truth about the limits of their knowledge." The Cronbach's alpha for the current study was .68.

2.4.1.2 *IPIP NEO-PI-R Agreeableness Trust facet*. The IPIP NEO-PI-R Agreeableness Trust facet [21] is a 10-item scale used to measure dispositional trust. Two items (i.e., "Suspect hidden motives in others" and "Am wary of others") were removed during data cleaning, due to these items reducing the internal consistency of the scale. A sample item from this scale is, "Trust others." The Cronbach's alpha for the current study was .83.

2.4.3. Dispositional Distrust. Dispositional distrust was assessed using seven items adapted from the General Distrust Scale [31]. Three items (i.e., "Most people can be trusted to behave fairly," "Most people can be trusted to cooperate with others," and "Most people can be trusted to behave selflessly") were removed, due to these items assessing dispositional trust. A sample item from this scale is, "Most people tend to be selfish." The Cronbach's alpha for the current study was .77.

2.4.4. Suspicion Propensity. Dispositional suspicion was assessed using seven items adapted from the Suspicion Propensity scale [27]. A sample item from this scale is, "I am naturally suspicious." The

Cronbach's alpha for the current study was .38. Note that due to the unreliability of this scale, we interpret the results below cautiously.

2.4.5. Trust. Context-dependent trust in the partner was assessed using 11 items adapted from the Trust in Automated Systems scale [32]. Participants completed this scale six times (i.e., once per round). This measure was chosen because the items were applicable to both partner conditions. The referent was modified from "the system" to "the runner." A sample item from this scale is, "The runner is dependable." Cronbach's alpha for the initial time point was .89.

2.4.6. Suspicion. Context-dependent suspicion was assessed using nine items adapted from [33]. Again, participants answered this scale once per each round. A sample item from this scale is, "I felt like I was being taken advantage of when doing this round." Cronbach's alpha for the initial time point was .85.

3. Results

We modeled the change in context-dependent trust and suspicion over time using a series of discontinuous growth models [34]. Discontinuous growth models are hierarchical linear models that incorporate an abrupt change(s) in the growth process, which aligns with the abrupt trust violations in the experimental design. Specifically, discontinuous growth models have three basic parameters to be coded including: (a) time (i.e., the initial linear slope before the abrupt change), (b) transition (i.e., the amount of change in the intercept after the specific event), and (c) the recovery (i.e., the amount of change in the linear slope after the event; see Table 1 for the coding strategies adopted). Note that the initial time point is coded as zero. In order to maintain consistency, we describe the time points from zero to five henceforth.

Researchers can use different coding techniques of the time variable to alter the interpretations of the time parameters [34]. Specifically, the initial time slope can be recoded to test whether the transition is significantly different from zero (i.e., absolute), rather than assessing the relative change in the intercept (see coding for Time.A in Table 1). In the current study, we used the absolute coding strategy for the time slope, which indicted whether or not the decrease in trust, and the increase in suspicion, was significantly different than zero. Finally, the time variable can also be coded to test the absolute change

in the recovery parameter (see the Time.R variable in Table 1). We used the absolute recovery term, which tests whether the slope occurring after the distrust behavior was significantly different from zero, rather than testing whether the recovery slope was significantly different from the initial time slope.

Our current model had two discrete transition periods (i.e., the distrust behavior that occurred at time point three and the trusting behavior that occurred at time point five; see Figure 1). Because the model with two transitions failed to converge, we omitted the second transition term and modeled only the first five measurements.

Table 1. Coding schemes for the discontinuous growth models

Time	Time.A	Time.R	Transition	Recovery
0	0	0	0	0
1	1	1	0	0
2	2	2	0	0
3	2	3	1	0
4	2	3	1	1

Note. All coding strategies were based on suggestions by Bliese & Lang (2016). Time.A depicts the absolute change in the transition event. Time.R depicts an absolute change in the recovery term.
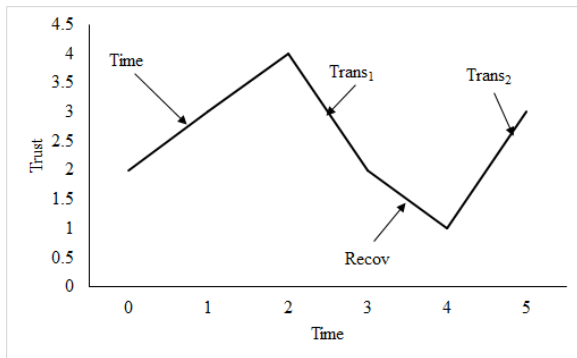

Figure 1. A hypothetical demonstration of the transition periods modeled in the current study. Note the fifth time point was removed from the statistical analysis, due to a lack of convergence.

In order to test these growth models, we followed a stepwise procedure outlined by Bliese and Lang [34] using the *nlme* package in R [35, 36]. First, the discontinuous growth model is an extension of the general growth model, in which a stepwise procedure is outlined by Bliese and Ployhart [37]. Generally, the first step frees the intercept, which allows all participants to have different initial levels of the outcome variable (i.e., trust, suspicion). Step 2 determines the appropriate change in the outcome variable over time (e.g., linear, quadratic) and simultaneously allows the participants to have different change patterns over time (i.e., random

slope variance). Discontinuous growth models focus on Step 2. Specifically, within the random slope variation, we coded where the transition (i.e., where the slope changes abruptly) and the recovery (i.e., the slope after the abrupt change) occurs. We placed the transition at the third time point (i.e., where the distrust behavior occurred), and we coded the recovery to occur at time point four (i.e., immediately after the transition). In Step 3, autocorrelation was tested to determine whether the correlations are stronger for time points occurring closer together in time, as well as testing the assumption of homoscedasticity. Finally, Step 4 involved adding person-level predictors of the growth pattern (e.g., dispositional trust). Following Bliese and Ployhart, we removed any term from the model that was non-significant.

In Step 4, we added the dispositional variables in a specific order, which was consistent for both trust and suspicion outcome models. The first model included two demographic variable (i.e., gender and age). We then tested whether dispositional trust, measured with the Propensity to Trust scale [19] and the IPIP NEO-PI-R Agreeableness Trust facet [21], predicted changes in trust over time simultaneously, due to both scales measuring the general tendency to trust others. Next, we added dispositional distrust to the model to test whether dispositional distrust predicted changes in outcomes after controlling for variance explained by the dispositional trust scale(s). We placed dispositional distrust into the second model, as research on dispositional distrust has been minimal. Finally, we added dispositional suspicion to the model in the final step, as this trait has received the least amount of research attention.

3.1. Trust in partner

The first model tested changes in perceived trust in partner over time. Note that Level-2 variables (e.g., partner condition) were not entered into the equation until Step 4, so Step 1 through Step 3 estimates were averaged across the partner conditions. In Step 1, we found an ICC estimate of .57, and thus we allowed initial trust perceptions to vary across participants (i.e., random intercept variance). In Step 2, we found evidence of a significant increase in trust throughout the first three rounds ($B = 0.27$, $t(193) = 3.89$, $p < .01$) and a significant decrease in trust after the distrust behavior in round 3 ($B = -1.06$, $t(193) = -8.00$, $p < .01$). The recovery term was also significantly different from zero, $B = -0.40$, $t(193) = -2.85$, $p < .01$, so we included the recovery term as a fixed effect. The slope variance for the time term, $\Delta\chi^2_{(2)} = 33.83$, $p <$

.01, and the transition term, $\Delta\chi^2_{(3)} = 39.12$, $p < .01$, were both significant, so we estimated a separate time and transition slope for each participant. We observed no significant slope variance for the recovery term, so we fixed the recovery term to be equal across all participants. Step 3 showed no significant autocorrelation of errors and no significant violations of homoscedasticity, and thus we omitted these from the model. Finally, we added the Level-2 variables into the model.

We found no significant differences for either the linear time variable or the transition period between the Partner types (i.e., robot versus human), so we removed the partner type from the model. Amongst the demographic variables (i.e., age and gender), we found no significant predictors for the linear or the transition periods. Next, we found that the Propensity to Trust scale [19] significantly moderated the transition phase and trust relationship while controlling for IPIP NEO-PI-R Agreeableness Trust facet scores, $\gamma = 0.54$, $t(189) = 2.07$, $p < .05$, see Table 2.

Specifically, the negative slope in trust following the distrust behavior was weaker for participants who scored higher on the Propensity to Trust scale (see Figure 2). IPIP NEO-PI-R Agreeableness Trust facet scores failed to predict changes in trust in the runner over time when controlling for scores from the Propensity to Trust scale. Dispositional distrust had no significant relationship with changes in trust over time. Finally, dispositional suspicion was a significant predictor of the intercept term, $\gamma = -0.49$, $t(46) = 2.00$, $p < .05$. This means that the initial levels of trust were lower for those who reported higher levels of dispositional suspicion.
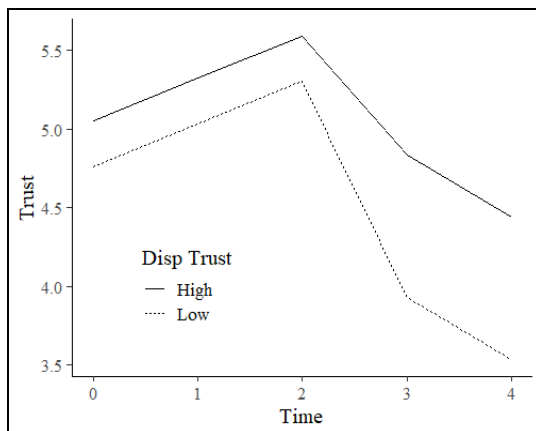

Figure 2. Differences in trust change over time depending on participants' level of dispositional trust measured by the Propensity to Trust Scale.

Table 2. Changes of trust over time with Level-2 predictors included.

Fixed Effect	Estimate	SE	df	t-value
Intercept	4.91	0.14	193	34.00**
Time	0.27	0.07	193	3.89*
Transition	-1.06	0.13	193	-8.00**
Recovery	-0.40	0.14	193	-2.85**
Partner[a]	-0.09	0.10	47	-0.86
Time*Partner[a]	0.03	0.16	191	0.68
Trans*Partner[a]	-0.17	0.13	191	-1.26
Age[b]	0.02	0.02	46	0.78
Gender[b]	-0.17	0.10	46	-1.65
Time*Age[b]	-0.01	0.01	189	-0.85
Time*Gender[b]	0.02	0.06	189	0.30
Trans*Age[b]	-0.04	0.03	189	-1.47
Trans*Gender[b]	-0.15	0.14	189	-1.06
PT[c]	0.24	0.20	46	1.19
IPTrust[c]	0.20	0.19	46	1.05
Time*PT[c]	0.13	0.12	189	1.04
Time*IPTrust[c]	-0.09	0.12	189	-0.77
Trans*PT[c]	0.54	0.26	189	2.07*
Trans*IPTrust[c]	-0.01	0.25	189	-0.02
Dis[d]	-0.17	0.18	46	-0.94
Time*Dis[d]	0.16	0.10	190	1.67
Trans* Dis[d]	0.20	0.24	190	0.85
SP[e]	-0.49	0.23	46	-2.09*
Time*SP[e]	0.13	0.14	190	0.94
Trans* SP[e]	0.01	0.32	190	0.03

Note. $N = 49$. Superscripts indicate the step in which the Level-2 predictors were entered. The recovery period was excluded, due to a lack of significant recovery variance across participants. Human was the reference group for Partner factor variable. Female was reference group for Gender. PT = Propensity to Trust scale. IPTrust = IPIP NEO-PI-R Agreeableness Trust facet scale. Dis = General Distrust scale. SP = Suspicion propensity scale. *$p < .05$. **$p < .01$.

3.2. Suspicion perceptions of partner

The second model tested the change in suspicion of the partner over time. In Step 1, we found an ICC estimate of .39, so we allowed each participant to have his/her own initial suspicion estimate. Again, Level-2 variables (e.g., partner condition) were not entered until Step 4, so Step 1 through Step 3 estimates were averaged across both partner conditions. In Step 2, we found evidence of significant increase in suspicion after the manipulation condition ($B = 1.04$, $t(193) = 6.32$, $p < .01$). The initial time slope term and the recovery term were not significantly different from zero. We observed significant slope variance for the linear time term, $\Delta\chi^2_{(2)} = 22.45$, $p < .01$, and the transition term, $\Delta\chi^2_{(3)} = 7.96$, $p < .05$. Thus, we estimated separate time and transition slopes for each participant. We found no significant autocorrelation of errors or violations of homoscedasticity, so we removed those

from the model. Finally, we added the Level-2 variables.

First, we tested whether any changes in suspicion over time could be predicted by whether the participant's partner was a human or robot. We found no significant differences in changes in suspicion over time across partner conditions, so we removed the partner type from the model. Amongst the demographic variables, we found that age was a positive, significant predictor of the increase in suspicion perceptions corresponding to the transition (i.e., immediately after the distrust behavior), $\gamma = 0.06$, $t(189) = 2.08$, $p < .05$. Thus, the increase in suspicion after the distrust behavior was stronger for older participants than for younger participants (see Figure 3). In the subsequent step, we found non-significant effects for dispositional trust using the Propensity to Trust scale [19], the IPIP NEO-PI_R Agreeableness Trust facet scale [21], nor the General Distrust scale [29], see Table 3.

Similar to the trust findings, we found that dispositional suspicion predicted participants' initial suspicion reports, $\gamma = 0.70$, $t(46) = 2.08$, $p < .05$. Thus, participants who reported higher levels of dispositional suspicion also reported higher levels of context-dependent suspicion at the first time point (i.e., time point 0). We recommend interpreting the dispositional propensity findings with caution, given the low reliability of the measure.
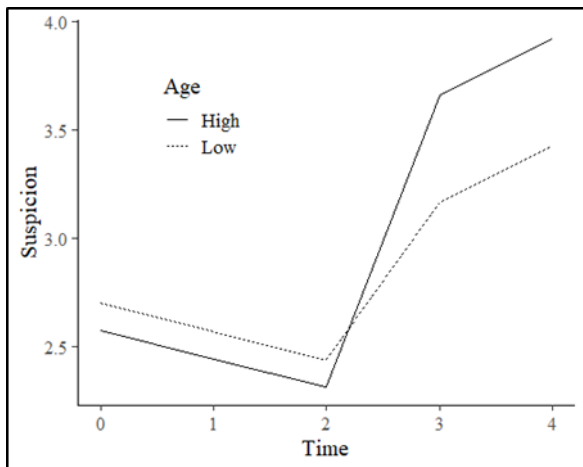

Figure 3. Differences in suspicion change over time depending on participants' age.

Table 3. Changes in suspicion over time with Level-2 predictors included.

Fixed Effect	Estimate	SE	df	t-value
Intercept	2.64	0.15	193	17.93**
Time	-0.13	0.09	193	-1.55
Transition	1.04	0.16	193	6.32**
Recovery	0.27	0.17	193	1.55
Partner[a]	0.09	0.15	47	0.59
Time*Partner[a]	-0.06	0.09	191	-0.65
Trans*Partner[a]	0.15	0.13	191	1.12
Age[b]	-0.04	0.03	46	-1.27
Gender[b]	0.14	0.15	46	0.92
Time*Age[b]	0.02	0.02	189	1.04
Time*Gender[b]	-0.03	0.09	189	-0.32
Trans*Age[b]	0.06	0.03	189	2.08*
Trans*Gender[b]	0.11	0.13	189	0.79
PT[c]	-0.23	0.29	45	-0.77
IPTrust[c]	-0.37	0.27	45	-1.35
Time*PT[c]	-0.12	0.18	188	-0.64
Time*IPTrust[c]	0.18	0.17	188	1.01
Trans*PT[c]	-0.30	0.26	188	-1.15
Trans*IPTrust[c]	0.15	0.24	188	0.62
Dis[d]	0.33	0.24	46	1.40
Time*Dis[d]	-0.11	0.15	190	-0.73
Trans* Dis [d]	-0.02	0.21	190	-0.12
SP[e]	0.70	0.33	46	2.08*
Time*SP[e]	-0.03	0.21	190	-0.16
Trans* SP[e]	0.23	0.30	190	0.75

Note. $N = 49$. Superscripts indicate the step in which the Level-2 predictors were entered. The recovery period term was excluded, due to a lack of significant recovery variance across participants. Human was the reference group for Partner factor variable. Female was reference group for Gender. PT = Propensity to Trust scale. Trust = IPIP NEO-PI-R Agreeableness Trust facet scale. Dis = General Distrust scale. SP = Suspicion propensity scale. *$p < .05$. **$p < .01$.

4. Discussion

In general, when participants' partners (i.e., either human or robot) engaged in behaviors detrimental to participants' goals (i.e., retuning the participant less money than promised), participants reported decreased trust and increased suspicion. Past research has shown others' distrust behaviors decrease trust [38], and we found similar findings using both human and anthropomorphized robot partners.

We observed no differences in trust or suspicion perceptions between a human partner and a robot partner, which deviates from prior findings that some people perceive intelligent technologies as being perfectly reliable (i.e., the perfect automation schema [8, 39]). If participants perceived the partner as infallible, we would have expected trust decrease to

be steeper after a distrusting behavior for the robot partner compared to the human partner. One explanation for these findings, however, is that context-dependent trust mediates interactions between both humans and robots [40]. If so, there may be no differences in perceived trust, regardless of whether the partner is a human or a robot. Alternatively, the backstory describing the Nao robot as having "self-preserving algorithms" may have attenuated the differences in human-human and human-robot trust by clearly stating a rationale for the robot to engage in a distrust behavior. Finally, note that prior research has questioned the degree to which interpersonal trust is appropriate with robot partners, as robots lack a sense of self [41]. Prior research, however, has shown that people attribute moral qualities to robots [42], which may have been highlighted by the robot's anthropomorphic features (see Method section above).

Overall, we found dispositional trust measured via the Propensity to Trust scale [19] significantly moderated the decrease in trust immediately following a distrust behavior. Specifically, those scoring higher in dispositional trust showed a less steep decrease in context-dependent trust after the distrust behavior compared to those lower in dispositional trust. Thus, it appears that high levels of dispositional trust may attenuate the effects trust violations have on participants' trust in a referent. Although we included only two rounds where the partner engaged in selfish behaviors, these findings suggest that dispositional trust may leave a person vulnerable to continuing to trust an entity that is less trustworthy than desired. Future research, however, should investigate whether this pattern is sustained with more observations of distrust behaviors.

The dispositional variables, however, failed to predict changes in suspicion over time. Age, however, significantly predicted the transition phase for suspicion. Specifically, older participants reported higher levels of suspicion after trust violations committed by their partners compared to younger participants. These results may suggest that older participants were more sensitive to their partners' behaviors than younger participants. One of the three characteristics of suspicion is increased cognitive activity [26, 27]. Future research should investigate whether older participants might be more willing to put forth the cognitive effort necessary to increase suspicion in the moment.

Implications from the current study involve the influence of both age and dispositional trust on changes in the trust process over time. First, age should be included as a covariate when researchers investigate changes in trust or suspicion over time,

especially if the experimental design includes a distrust behavior. If researchers fail to consider the effects of age on suspicion in the moment, they are omitting an important explanatory factor throughout the trust process.

Practitioners should consider how a person's dispositional trust may affect his or her trust of either human or robot partners. There are two possible practical arguments for these findings. First, a trustor should see the trustee as less trustworthy when the trustee engages in actions that undermine the trustor's well-being and practical goals. If dispositional trust undermines the extent to which a person can accurately monitor the trustworthiness of a partner that engages in selfish acts, researchers and practitioners alike should be very clear in explaining to people—especially those who have high dispositional trust tendencies—the negative effects of unwarranted trustworthiness perceptions to attenuate improperly calibrated trust towards the referent.

On the other hand, propensity to trust may limit the degree to which people prematurely lose trust in a partner who makes one or two mistakes. As previous research has shown that violations of trust can lead to automation disuse (i.e., refusing to use automated aids that would help performance; see [43]), it is important that trust is not abandoned prematurely. More research is needed to further delineate appropriate levels of trust over time given the person's level of dispositional trust.

There were limitations to this research. First, our study had limitations associated with the sample. The sample included only undergraduate students. Thus, we may observe different results if we conducted the study with a different age group. For example, previous research has shown that trust in automation is higher for older adults compared to younger adults [44, 45]. Note, however, that we observed differences in suspicion changes across time with a sample consisting of mostly young adults. Future research should study whether the results can be replicated with a sample that includes older adults. The sample size may have also been small for the DCM analyses, especially with the few number of observations within subjects. Thus, future research should expand on these findings with larger sample sizes.

Prior research has suggested that trust differences may occur across cultures [46, 47]. Because our sample included undergraduate students from a Midwestern American university, these relationships should be tested using samples from different countries of origins and ethnicities. Finally, the dispositional suspicion measure showed an unacceptably low internal consistency estimate. Although prior research has shown the Suspicion

Propensity scale [27] to have multiple facets, which may attenuate observed internal consistency estimates, future research should investigate further the psychometric properties of this scale. Participants may have also found it difficult to separate dispositional distrust and suspicion propensity, as these variables show similar theoretical overlap. The observed correlation between the two individual difference variables was .56, which provides some evidence that, although related, participants were able to distinguish between the two variables.

5. Acknowledgements

6. References

[1] S.L. Norris, and R.L. Zweigenhaft, "Self-monitoring, Trust, and Commitment in Romantic Relationships", The Journal of Social Psychology, Routledge, United Kingdom, 1999, pp 215-220.

[2] Lewicki, R.J., and B.B. Bunker, Developing and Maintaining Trust in Work Relationships, Trust in Organizations: Frontiers of Theory and Research, SAGE Publications, Inc., United States, 1996.

[3] R.C. Mayer, J.H. Davis, and F.D. Schoorman, "An Integrative Model of Organizational Trust", Academy of Management Review, Academy of Management, United States, 1995, pp 709-734.

[4] D.H. McKnight, L.L. Cummings, and N.L. Chervany, "Initial Trust Formation in New Organizational Relationships", Academy of Management Review, Academy of Management, United States, 1998, pp 473-490.

[5] J.A. Colquitt, B.A. Scott, and J.A. LePine, "Trust, Trustworthiness, and Trust Propensity: A Meta-analytic Test of their Unique Relationships with Risk Taking and Job Performance", Journal of Applied Psychology, American Psychological Association, United States, 2007, pp 909-927.

[6] K.A. Hoff, and M. Bashir, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust", Human Factors, SAGE Publications Inc., United States, 2015, pp 407-434.

[7] J.D. Lee, and K.A. See, "Trust in Automation: Designing for Appropriate Reliance", Human Factors, SAGE Publications Inc., United States, 2004, pp 50-80.

[8] S.M. Merritt, H. Heimbaugh, J. LaChapell, and D. Lee, "I Trust it, but I Don't Know Why: Effects of Implicit Attitudes toward Automation on Trust in an Automated System", Human Factors, SAGE Publications, Inc., United States, 2013, pp 520-534.

[9] G. Abe, M. Itor, & K. Tanaka, "Dynamics Of Drivers' Trust in Warning Systems", IFAC, Elsevier, Netherlands, 2002, pp 363-368.

[10] S. Hergeth, L. Lorenz, R. Vilimek, and J.F. Krems, "Keep your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust during Highly Automated Driving", Human factors, SAGE Publications Inc., United States, 2016, pp 509-519.

[11] N.R. Bailey, and M.W. Scerbo, "Automation-induced Complacency for Monitoring Highly Reliable Systems: The Role of Task Complexity, System Experience, and Operator Trust", Theoretical Issues in Ergonomics Science, Taylor & Francis, United Kingdom, 2017, pp 321-348.

[12] E. de Visser, and R. Parasuraman, "Adaptive Aiding of Human-Robot Teaming: Effects of Imperfect Automation on Performance, Trust, and Workload", Journal of Cognitive Engineering and Decision Making, SAGE Publications Inc., United States, 2011, pp 209-231.

[13] P. Madhavan, and D.A. Wiegmann, "Similarities and Differences between Human–Human and Human–Automation Trust: an Integrative Review", Theoretical Issues in Ergonomics Science, Taylor & Francis, United Kingdom, 2007, pp 277-301.

[14] S.M. Merritt, and D.R. Illgen, "Not All Trust is Created Equal: Dispositional and History-Based Trust in Human-Automation Interactions", Human Factors, SAGE Publications Inc., United States, 2008, pp 194-210.

[15] G.M. Alarcon, J.B. Lyons, and J.C. Christensen, "The Effect of Propensity to Trust and Familiarity on Perceptions of Trustworthiness Over Time", Personality and Individual Differences, Elsevier, Netherlands, 2016, pp 309-315.

[16] G.A. Bigley, and J.L. Pearce, "Straining for Shared Meaning in Organization Science: Problems of Trust and Distrust", The Academy of Management Review, Academy of Management, United States, 1998, pp 405-421.

[17] D.H. McKnight, L.L. Cummings, and N.L. Chervany, "Initial Trust Formation in New Organizational Relationships", Academy of Management Review, Academy of Management, United States, 1998, pp 473-490.

[18] J.B. Rotter, "A New Scale for the Measurement of Interpersonal Trust" Journal of Personality, American Psychological Association, United States, 1967, pp 651-665.

[19] R.C. Mayer, and J.H. Davis, "The Effect of the Performance Appraisal System on Trust for Management: A Field Quasi-experiment", Journal of Applied Psychology, American Psychological Association, United States, 1999, pp 123-136.

[20] G.M. Alarcon, J.B. Lyons, and J.C. Christensen, M.A. Bowers, S.L. Klosterman, and A. Capiola, "The Role of

Propensity to Trust and the Five Factor Model Across the Trust Process", Journal of Research in Personality, Elsevier, Netherlands, 2018, pp 69-82.

[21] L.R. Goldberg, "A Broad-Bandwidth, Public Domain, Personality Inventory Measuring the Lower-level Facets of Several Five-factor Models", Personality Psychology in Europe, John Wiley & Sons, United States, 1999, pp 7-28.

[22] S. Jessup, T. Schneider, G. Alarcon, T. Ryan, and A. Capiola, "The measurement of the propensity to trust automation", Human-Computer Interaction Annual Conference, United States, in press.

[23] Y.S. Chang, and S.R. Fang, "Antecedents and Distinctions between Online Trust and Distrust: Predicting High-and Low-risk Internet Behaviors", Journal of Electronic Commerce Research, California State University Press, United States, 2013, pp 149-166.

[24] F.D. Schoorman, R.C. Mayer, and J.H. Davis, "An Integrative Model of Organizational Trust: Past, Present, and Future", Academy of Management Review, Academy of Management, United States, 2007, pp 344-354.

[25] R.J. Lewicki, D.J. McAllister, and R.J. Bies, "Trust and Distrust: New Relationships and Realities", Academy of Management Review, Academy of Management, United States, 1998, pp 438-458.

[26] P. Bobko, A.J. Barelka, and L.M. Hirshfield, "The Construct of State-level Suspicion: A Model and Research Agenda for Automated and Information Technology (IT) Contexts", Human Factors, SAGE Publications, Inc., United States, 2014, pp 489-508.

[27] C. Calhoun, P. Bobko, M. Schuelke, S. Jessup, T. Ryan, C. Walter…C. Stokes. "Suspicion, trust, and automation." SRA International Inc. Publication No. AFRL-RH-WP-TR-2017-0002, 2017.

[28] A. Capiola, S. Jessup, T. Ryan, and G. M. Alarcon, "Exploring the Unique and Shared Variance of Propensity to Trust and Suspicion Propensity", Journal of Individual Differences, Hogrefe Publishing Corp., United States, in press.

[29] G.M. Alarcon, J.B. Lyons, J.C. Christensen, S.L Klosterman, M.A. Bowers, T.J. Ryan, and K.T. Wynne, "The effect of propensity to trust and perceptions of trustworthiness on trust behaviors in dyads". Behavior Research Methods, Springer International, Germany, 2018, pp 1906-1920.

[30] J. Berg, J. Dickhaut, K. McCabe, "Trust, Reciprocity, and Social History", Games and Economic Behavior, Elsevier, United States, 1995, pp 122-142.

[31] M. Mooijman, W.W. van Dijk, N. Ellemers, and E. van Dijk, "Why Leaders Punish: A Power Perspective", Journal of Personality and Social Psychology, American Psychological Association, United States, 2015, pp 75-89.

[32] J.Y. Jian, A.M Bisantz, and C.G Drury, "Foundations for an Empirically Determined Scale of Trust in Automated Systems", International Journal of Cognitive Ergonomics, Routledge, United Kingdom, 2000, pp 53-71.

[33] P. Bobko, A.J. Barelka, L.M. Hirshfield, and J.B. Lyons, "Invited Article: The Construct of Suspicion and how it can Benefit Theories and Models in Organizational Science", Journal of Business and Psychology, 2014, pp 335-342.

[34] P.D. Bliese, and J.W. Lang, "Understanding Relative and Absolute Change in Discontinuous Growth Models: Coding Alternatives and Implications for Hypothesis Testing" Organizational Research Methods, SAGE Publications, Inc., United States, 2016, pp 562-592.

[35] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R.C. Team, "nlme Linear and Nonlinear Mixed Effects Models" R Package Version, 2013, p 111.

[36] R Core Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, 2019.

[37] P.D. Bliese, and R.E. Ployhart, "Growth Modeling Using Random Coefficient Models: Model Building, Testing, and Illustrations", Organizational Research Methods, SAGE Publications, Inc., United States, 2002, pp 362-387.

[38] S.L. Robinson, "Trust and Breach of the Psychological Contract", Administrative Science Quarterly, SAGE Publications, Inc., United States, 1996, pp 574-599.

[39] M.T. Dzindolet, L.G. Pierce, H.P. Beck, and L.A. Dawe, "The Perceived Utility of Human and Automated Aids in a Visual Detection Task", Human Factors, SAGE Publications, Inc., United States, 2002, pp 79-94.

[40] T.B. Sheridan, "Considerations in modeling the Human Supervisory Controller", IFAC Proceedings Volumes, Elsevier, Netherlands, 1975, pp 223-228.

[41] B. Friedman, P. Kahn, & J. Hagman, "Trust online." Communications of the ACM, United States, 2000.

[42] B. Friedman, P. Kahn, & J. Hagman, "Hardware companions?: What online AIBO discussion forums reveal about the human-robotic relationship", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, United States, 2003.

[43] J. D. Lee, and N. Moray, "Trust, Self-confidence, and Operators' Adaptation to Automation", International Journal of Human-computer Studies, Elsevier, Netherlands, 1994, pp 153-184.

[44] B. Donmez, L.N. Boyle, J.D. Lee, and D.V. McGehee, "Drivers' Attitudes toward Imperfect Distraction Mitigation Strategies", Transportation Research part F: traffic psychology and behaviour, Elsevier, Netherlands, 2006, pp 387-398.

[45] G. Ho, D. Wheatley, and C.T. Scialfa, "Age Differences in Trust and Reliance of a Medication Management System", Interacting with Computers, Oxford Academic, United States, 2005, pp 690-710.

[46] E. Huerta, T. Glandon, and Y. Petrides, "Framing, Decision-aid Systems, and Culture: Exploring Influences on Fraud Investigations", International Journal of Accounting Information Systems, Elsevier, Netherlands, 2012, pp 316-333.

[47] S. Rice, K. Kraemer, S.R. Winter, R. Mehta, V. Dunbar, T.G. Rosser, and J.C. Moore, "Passengers From India and the United States have Differential Opinions about Autonomous Auto-pilots for Commercial Flights", International Journal of Aviation, Aeronautics, and Aerospace, Taylor & Francis, United Kingdom, 2014, pp 3-12.