Building Trust Over Time in Human-Agent Relationships

Danielle Marie Olson*

dolson@mit.edu

MIT Computer Science and Artificial Intelligence

Laboratory

Cambridge, Massachusetts, USA

Yan Xu

yanx@fb.com

Facebook Reality Labs Research

Redmond, WA, USA

ABSTRACT

This paper aims to understand how long-term trust and distrust develop between humans and agents (smart objects). We first conducted a qualitative study to explore key factors that lead to trust and distrust, how the human-agent trust journey develops, and what roles these trust-building factors play in the journey. This qualitative study involved an open-ended questionnaire completed by 621 participants and a five-day diary study with 60 participants that examined 499 human-object and human-agent relationships. Next, we conducted a mixed methods study with 146 participants to rate the importance of key factors identified through analysis of the data collected in the first study. We contribute to the HAI community by identifying eight factors accounting for participants' trust and distrust toward new objects, how these factors play different roles at different phases of the human-agent trust journey, and how important these factors are for both trust and distrust. We identified ebbs and flows in the human-agent trust journey over time, revealing time periods when trust is particularly vulnerable. We also found the most important factors for building trust and avoiding distrust did not entirely overlap. We discuss these findings and their implications for designing agents which need to foster trustful long-term relationships with humans.

CCS CONCEPTS

• Human-centered computing → User studies; Field studies.

KEYWORDS

human-AI; human-agent interaction; trust; artificial intelligence; smart devices; diary study

ACM Reference Format:

Danielle Marie Olson and Yan Xu. 2021. Building Trust Over Time in Human-Agent Relationships. In *Proceedings of the 9th International Conference on Human-Agent Interaction (HAI '21)*, November 9–11, 2021, Nagoya, Japan. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3472307.3484178>

*Currently at Apple Inc.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs International 4.0 License.

HAI '21, November 9–11, 2021, Nagoya, Japan
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8620-3/21/11.
<https://doi.org/10.1145/3472307.3484178>

1 INTRODUCTION

The rapid growth of technologies such as artificial intelligence, extended reality, and ubiquitous computing have introduced numerous sensor-equipped agents (e.g., smart devices and assistants) to humans' everyday lives. These agents have become more autonomous and intelligent; they are no longer passive and interaction-specific tools. Rather, their “eyes, ears, and brains” continuously observe, compute, and optimize even when people are not explicitly interacting with them. As a result, human-agent interaction (HAI) is gravitating from a transactional model based on repeated short-term interactions toward a long-term partnership between humans and intelligent agents [8]. Trust is a critical factor influencing success or failure of these long-term partnerships. As the number of these relationships continue to multiply, HAI researchers and designers can benefit from understanding how humans develop, evaluate, and maintain longstanding trust with objects they interact with over long periods of time.

Prior work on trust in HAI has generally examined how individual traits, agent performance, and situational factors influence trust/distrust in short-term and experimental settings. To gain insights about long-term development and evolution of trust which may otherwise remain latent in controlled (or laboratory) settings, we wanted to study long-term trust as it develops within existing relationships in humans' everyday lives. To address this current gap in the literature, we investigated the following research question: **How do long-term trust and distrust develop in human-object and human-agent relationships?** In addition to this primary research question, we also collected insights on these sub-questions:

- What are the factors that influence trust and distrust during different stages of time in the relationship?
- What is the relative importance of these factors in terms of how they impact overall trust and distrust?

To answer these questions, we conducted a three-phase user study. First, we started with a broad examination of the trust landscape by asking 621 participants about mechanisms by which objects earned their trust/distrust. We synthesized this data to identify a list of the most influential factors leading to trust/distrust in long-term human-object relationships. Second, we collected rich, detailed data about 499 trustful/distrustful human-object relationships by conducting a five-day diary study with 60 participants. We did not limit all diary entries to be about agents (or smart objects). Rather, our intention was to learn about the broad domain of long-term human-object relationships (which included, but was not limited to, relationships with agents). Third, we ranked the relative importance of trust/distrust-building factors through a mixed methods study with an independent group of 146 participants. In addition to empirical findings from our studies, we contribute a new methodological

approach to the HAI community by (1) expanding the scope of focus solely from agents to also include objects, (2) using a diary study to uncover context-rich insights in people's everyday life settings with objects they already own/use, and (3) using a study protocol explicitly framing HAI over time as a relationship with humans. In this paper, we report on these novel contributions and conclude by presenting a set of evidence-based design recommendations to designing agents which foster trust in long-term relationships with humans.

2 BACKGROUND

We define an *agent* as an autonomous or semi-autonomous computer system which has four key properties: autonomy, reactivity, proactiveness, and social ability [14]. In this model, an agent's purpose is to notice potentially relevant events, recognize those events, and act on them by applying a set of rules [7]. We consider specific types of agents to exist along a spectrum, from more passive assistants (e.g., ubiquitous voice assistants which are triggered using wake words), to more proactive and autonomous partners (e.g., co-robots who anticipate actions of human collaborators and assist with tasks in a safe and compliant way) [2]. Our reference to agents throughout this work includes smart devices, assistants, and other kinds of systems which have these properties but may differ in their points along this spectrum (from passive to autonomous).

We applied three frameworks in our analysis of human-agent trust throughout our investigation. First, as described in [13], human-agent trust can be broken down into two components: (1) a cognition-based component (linked to the perceived understandability and technical competence of the system), and (2) an affect-based component (linked to faith in and personal attachment to the system). Second, we focused on three dimensions influencing trust in agents described in [12]. These bases of trust are linked to having confident and positive expectations about the agent regarding its (1) performance (the agent's stability, consistency, reliability and predictability), (2) process (attributions and inferences made based on an agent's operation and behavior), and (3) purpose (the perception of the designer's intent and why the agent was developed). Finally, we examined trust through the lens of three interdependent layers of variability described in [10]: (1) dispositional trust (influenced by an individual's personality and tendency to trust agents), (2) situational trust (influenced by the internal and external variables within a given context), and (3) learned trust (influenced by direct interaction with the agent). There are two types of learned trust: initial learned trust (initial trust in an agent based on reputation, relevant past experiences, or knowledge of a system) and dynamic learned trust (trust in an agent as a result of direct interaction). The design of an agent has the most influence and control over dynamic learned trust as it is formed as a result of the human-agent interaction process over time. These frameworks guided the development of the assessment tools designed for this study as well as our data analysis.

Much prior work has generally examined how human traits, agent performance, and situational factors influence trust/distrust over short periods of time, with experimental agents (i.e., agents not already owned/used by participants), and in controlled (or laboratory) settings. However, less work has been conducted on how

trust develops within existing human-agent relationships in people's natural contexts over long periods of time. A 2020 systematic literature review on trust in smart personal assistants identified this gap as an important opportunity for HAI researchers to address: "From a methodological perspective, as most of the identified studies are based on laboratory experiments, we recommend to increasingly use field experiments to ensure external validity and, thus, to be able to provide stronger insights for practitioners" [15]. This paper contributes a novel methodological approach to support this research agenda. Additionally, to the best of the authors' knowledge, it is the first study to apply the study of long-term, existing human-object relationships to the domain of HAI.

3 METHODOLOGY

3.1 Study Design

We designed a multi-phase study to first qualitatively understand how trust/distrust evolve in people's long-term relationships with objects, including agents; what factors are playing critical roles at different phase of trust journey. Then, we sought to understand the relative importance of these factors. Our study design was structured as follows:

(1) Qualitative studies:

- Open-ended questionnaire (621 participants): Collected open-ended data about most important factors influencing participants' trust/distrust in objects
- Diary study (60 participants) and moderated interviews (9 participants): Collected diary entries over five-day period about trusted/distrusted objects and agents participants have owned/used, followed by in-depth interviews with a subset of participants

(2) Mixed methods study: Survey (146 participants; different participant pool than the qualitative studies): Collected participants' rankings of the top three most important factors influencing their trust/distrust in agents they have owned/used

3.2 Qualitative Studies

3.2.1 Open-ended Questionnaire. To investigate how people develop trust/distrust toward new objects, we conducted a study with an open-ended questionnaire with 621 participants recruited on the dScout user study platform [6]. The participants were asked to answer the following questions via a video or text entry:

- What does it take for you to feel fully trusting in a new object (whether it be technology or something else)? Why?
- What are the characteristics of a new object (whether it be technology or something else) that make you distrustful? Why?

We also collected participants' self-reported comfort level with new technologies, tendencies to trust or be cautious toward new technologies, app and device usage, and self-reported personality traits (i.e., openness, extroversion, and agreeableness [9]). We used this data to ensure diversity across these traits for the group of participants invited to the diary and interview study. A total of 621 participants completed the first survey study. All were located in the U.S. 58.8% of participants were women, 39.5% were men, 0.8% were non-binary, and 0.8% did not provide this information.

Participants' ages ranged from 18 to 70 years old ($M = 34.5$, $SD = 12.2$). 55.9% of participants were white, 14.7% were Black or African American, 12.2% were Asian, 10.6% were Hispanic or Latinx, 2.8% did not provide this information, 1.9% preferred to self-identify, 1.0% were American Indian or Alaska Native, 0.8% were Middle Eastern or North African, and 0.2% were Native Hawaiian or other Pacific Islander.

3.2.2 Diary Study and Moderated Interviews. To investigate how participants developed, evaluated, and maintained trust/distrust in long-term human-object and human-agent relationships in their everyday lives, we conducted a diary study on dScout. We sought to gain insight into participants' existing human-object and human-agent relationships within their everyday spatial, social, and temporal contexts. For this reason, we employed the diary study research method for their "high ecological value as they are carried out *in situ*, in the users' real environments" [5]. We selected 60 participants from the 621 participants who completed the open-ended questionnaire, balancing for diversity in terms of demographics, self-reported dispositional trust, comfort levels with new technologies, and personality traits. 55% of participants were women and 45% were men. Participants' ages ranged from 18 to 68 years old ($M = 35.9$, $SD = 12.0$). 36.7% of participants were white, 20% were Asian, 18.3% were Black or African American, 11.7% were Hispanic or Latinx, 8.3% preferred to self-identify, and 5.0% did not provide this information. The diary study lasted for five days (from a Wednesday to a Sunday to include both weekday and weekend interactions) and was structured as follows:

- **Diary entries (days 1–4):** Each day, participants logged information about one trusted object and one distrusted object. They were encouraged to take a video interacting with and discussing the object and were required to log object meta-data (e.g., Likert items scoring their overall trust/distrust in the object, duration of ownership/usage, whether the object is technology-related or not, and whether they consider the object "smart" or not).
- **Love/Break-up letters (day 5):** On the fifth (final) day of the diary study, the participants were asked to write a love letter toward one of their trusted technology-related objects, and a break-up letter toward one of their distrusted technology-related objects. Both objects needed to be owned/used for 6 months or longer. We prompted participants to write these letters with attention to the temporal dimension of the relationship development (i.e., how trust/distrust evolved over time). Participants were required to record a two-minute video of themselves reading each letter out loud, as if they were speaking directly to the object.
- **Moderated interviews:** We interviewed a subset of 9 participants (among the 60 who completed the diary study) to refine the emergent qualitative themes and gather more fine-grained details about micro-interactions during the relationship over time. Participants were strategically identified to provide a diverse representation of the overall diary study population. We also included two role-playing activities where participants personified the trusted and distrusted technologies that were the subjects of their love/breakup

letters and engaged in discussions about the technologies' characteristics and their feelings about them.

3.2.3 Qualitative Data Analysis. We conducted grounded theory open coding and memo writing [3] on the collected qualitative data. Our goal was to identify emergent themes regarding how trust/distrust evolved during different stages of time in the relationships from the interview and diary study transcripts. We also performed word cloud analysis to analyze the frequency of the descriptive words from participants' transcripts to ensure these trends were captured by our manual thematic analysis.

3.3 Mixed Methods Study

To understand the relative importance of the trust/distrust-building factors we identified from the qualitative study, we conducted a survey with an independent group of participants (i.e., no participants from the qualitative studies were eligible to participate). We conducted a survey with 146 participants on dScout asking them to name one smart device (or agent) they trust, and select the three most important reasons for their trust. They were also asked to name one smart device (or agent) they distrust, and select the three most important reasons for their distrust. The list of options for this survey was generated from the themes identified through the open-ended questionnaire qualitative data analysis, with an open-ended option (i.e., "other") to enable participants to specify any reasons not in the list.

4 RESULTS AND FINDINGS

4.1 Emergent Qualitative Themes on Influential Factors to Trust and Distrust

We identified eight themes through grounded theory coding of the qualitative data on what participants described as influencing their feelings of trust/distrust toward new objects and technologies. This section presents each of these themes in alphabetical order, which are labeled, described, and accompanied by representative participant quotes below, with key implications of these findings for HAI designers.

4.1.1 Brand or manufacturer. Participants were less likely to trust a new product or service if it was developed by a less well-known company or organization. However, they also had strong attitudes and opinions about the perceived trustworthiness of large organizations (e.g., Apple, Google, Facebook, Amazon, Dell, HP, Samsung) and used them as examples to demonstrate their concerns and values with regards to trust. A significant number of participants mentioned the perceived ethics and social values of a company and/or its leadership as influencing their brand trust/distrust (e.g., poor ethical practices resulted in more distrust in the brand).

"I'm generally trusting but if there is already some bad news about the company's privacy or about the company's habits I'm less likely to be trusting." –Woman, 22 years old

It is critical to consider how public perceptions of a brand's social values and ethics will affect consumers' willingness to adopt new products and services in the future. It is worthwhile to explore best

practices of companies that have earned and sustained long-term consumer trust.

4.1.2 Product performance (accuracy, precision, reliability). Participants described specific examples of “tests” and “trials” they put new objects through to evaluate and assess the performance of a new object. The results of these activities significantly impacted their trust/distrust. They described the initial tests as being lower stakes. If they perceived positive results from the initial tests, they increased their expectations over time, developing trust incrementally based on their perceptions of its accuracy, precision, reliability, and other performance-related traits.

“To feel fully trusting in a new object or perhaps technology I would have to have it used under rigorous conditions and for a while . . . I was getting a new headlamp to climb Mount Shasta early in the morning. It had to be rigorous to withstand cold temperatures and something that couldn’t have gone out in the middle of a climb in the dark. Obviously for safety purposes. So I had to vet this headlamp for a while and use it for a bit and also subjected it to some more rigorous conditions.” –Man, 51 years old

It is critical for a product or service to master (or even excel at) the initial, basic test cases consumers expect will be performed well during the out-of-box experience. Although consumers don’t expect the technology to be perfect, their trust will be significantly compromised if an object is unable to perform well during initial test cases or consistently perform well with increasingly demanding test cases over time.

4.1.3 Product quality (materials, durability, aesthetics). Participants expressed the perceived quality of an object influenced their perceptions of its trustworthiness. Factors such as the materials, anticipated durability, and aesthetics influenced judgments of quality. In addition to directly reviewing these object properties, an unexpected finding related to this theme was precise pricing being frequently mentioned as influencing judgments of object quality.

“I’ve bought things in the past for a very low price and they ended up breaking or not working properly. Price usually indicates quality, so it’s important that an object has a reasonable price if it’s supposed to be good quality.” –Woman, 20 years old

Pricing must strike a difficult balance between affordability and perceived value and quality. Pricing too low will result in consumers being suspicious of low quality, and pricing too high will create a higher barrier to entry to consumers who may be willing to try or buy a new product or service.

4.1.4 Reviews or word of mouth referrals. The most common theme described by participants as impacting their trust/distrust in an object was conducting research online by reading and watching reviews. This was most commonly done through crowdsourced reviews (e.g., Amazon, Reddit, Yelp) but a significant number of participants described relying on ratings and reviews from subject matter experts and trusted organizations (e.g., CNET, Better Business Bureau). Participants also relied on trusted social relationships

(e.g., friends, family, coworkers) and influencers (e.g., YouTube creators). If someone they knew expressed trust in a product or service (particularly if the individual trusted them for their knowledge and expertise in a type of product, such as technology), they were much more likely to trust the product, as well.

“I usually do a lot of research and reading [of] people’s reviews. If the technology has had positive reviews and a lot of good feedback I generally go into using that technology with a high level of trust to start with.”
–Man, 31 years old

The majority of consumers will not even consider trying or buying a new product if ratings and reviews are negative. Therefore, first impressions for early adopters who are likely to review and rate a new product are incredibly important. Gaining the trust of one individual can have a strong “domino effect” on gaining the trust of others, particularly if they have a large social network. Similarly, losing one individual’s trust can also have a strong ripple effect, deteriorating the potential for a product or service to be trusted by individuals in their social network. Thus, gaining the trust of highly-connected individuals could be a key strategy in the success of a new product or service launch.

4.1.5 Security and privacy. Participants expressed significant concerns about companies’ security and privacy practices, citing concerns about the collection of their personal information and their “digital footprint.” They expressed a desire to be able to use products and services which would bring as much value to their lives as possible with as little of their personal information as possible due to perceived security and privacy risks. However, in spite of expressing distrust of certain objects due to perceived negative security and privacy practices, several participants still used these objects due to feeling there was no comparable alternative.

“I feel I’ve been burned in the past with technology. So these days it’s going to be a situation where they have to earn my trust. I think about companies like [technology companies]. They aren’t trustworthy. They’re incredibly biased. They sell your information. They don’t protect your privacy. They’re just rotten. They really are and I still have to use them because sometimes, like [technology company], people use that for get togethers and stuff.” –Woman, 53 years old

Granular privacy controls were preferred for granting consent to apps and services to use personal information. Lengthy privacy statements resulted in more expressed distrust and concerns. Many participants expressed the desire for products and services to be clear and direct about exactly why they needed personal information and how it would be used. Finally, several participants described apps’ repeated requests for granting additional permissions as “desperate” and expressed their distrust toward this approach. Concerns in security and privacy lead to perceived untrustworthiness.

4.1.6 Simplicity or transparency (understanding how it works). Participants expressed simplicity of the instructions and user experience of a product or service not only increased their trust in the object, but also their trust in themselves to properly use and gain

experience with the object. Participants described how the use of familiar and conventional interaction patterns in new objects helped to grow their trust.

“What I’m looking for is something that I fully trust. I look for three things: simplicity, transparency, and a benefit for myself. Number one, I want a product or piece of technology that’s easy to use. If it’s simple to use, it’s not trying to give me the runaround and make me do something that I am unaware of.” –Man, 30 years old

Keeping instructions and user experience as simple, concise, and familiar as possible is critical for developing trust in a wide range of consumers spanning different levels of technology savvy. Furthermore, ensuring instructions are easily interpreted and documentation appears with a simple search engine query) is very important to consumers. Finally, there is a link between participants’ trust in an object and trust in themselves. The better accompanied an object is by documentation and support resources, the more individuals will trust themselves to use it, and thus their trust in an object will grow.

4.1.7 Time and personal experience. Trust was described as being developed in “levels” over time and through personal experience. Regardless of participants’ dispositions to trust/distrust new objects or their initial optimism/suspicion, trust was gained and lost, assessed and re-assessed through a relationship-building process taking months or years, depending on the specific context.

“So trust, for me, is something that does have to be built over time. I think that if I were to categorize myself, I initially trust things in the beginning but I earn more trust as time progresses. And that comes through a succession of many positive experiences with either a person or, in this case, like you’re talking about, an object. When I have a sort of consistent pattern of positive experiences, then I see my trust level going up.” –Woman, 45 years old

Even if participants continued to keep or use an object for a long period of time, it did not necessarily mean they trusted the object. Thus, time and experience are not necessarily positively correlated with trust. It is important to develop specific metrics and evaluation criteria for consumer trust, rather than rely on time or usage to indicate a consumer’s trust in the product or service. Many participants mentioned they continue to keep or use products and services they deem untrustworthy simply due to barriers preventing them from replacing or transitioning away from it. For example, some participants shared they kept home appliances they did not trust to function reliably simply because the cost of replacement was too high. Earning (and repairing) trust should remain a critical priority for designers regardless of how long people have owned/used a system.

4.1.8 Warranty or customer support. Participants’ trust in a brand’s (or manufacturer’s) warranty or customer support practices encouraged them to take more “risks” with trying, buying, and putting trust in new objects. Participants mentioned feeling more trust toward the product if the process for processing a warranty or

Figure 1: A selection of images submitted by 60 diary study participants. 499 trusted and distrusted objects were logged over a period of 5 days.

reaching out to customer support was straightforward and efficient.

“The characteristics of a new object that makes me distrustful would be if there isn’t a warranty or active customer service that I can utilize... I often try to plan for the worst-case scenario and that involves replacing an object or reaching out to customer service for support.” –Woman, 31 years old

Responsive warranty and customer support services serve as an insulating factor to taking a “risk” by trying a new product or service. Participants also described specific electronic commerce platforms as trustworthy, which in turn developed their trust in the objects they ordered using it.

4.2 Trust Development Journey Over Time

Our diary study resulted in a novel dataset of 499 objects with structured metadata. 51.7% of objects (258) were classified as trusted and 48.3% (241) as distrusted. The time owned/used for the objects in the dataset ranged between a few weeks to more than 20 years. Figure 1 shows a selection of images and screenshots of objects from this dataset. The majority of objects were technology-related (52.9% of total objects, or 264). Of the technology-related objects submitted, the majority of these were classified as intelligent (65.9% of technology-related objects, or 174). The number of trusted and distrusted objects across four categories based on a combination of technology status (i.e., technology-related or not) and intelligence status (i.e., smart object or not) was roughly equal (difference between number of trusted and distrusted objects per category: $M = \pm 4.3$, $SD = 11.4$). Through our qualitative analysis of the diary study, we identified patterns of micro-interactions participants described as influencing their trust/distrust at different stages of time. The goal of this analysis was to develop a descriptive (not predictive) model of human-object trust/distrust over time. The interactions and progressions between stages varied with a significant number of internal and external factors as indicated by participants’ diverse and numerous examples. The major takeaway is the temporal dimension of the human-object relationship influenced which

Table 1: Emergent themes from qualitative analysis of the diary study and moderated interview data on how trust and distrust are developed, evaluated, and maintained at different stages of time.

Time Period	Themes
Pre-Acquisition	<p>Indirect knowledge: Trust development begins before an individual owns/uses the object. Indirect knowledge from external sources influence initial trust/distrust of the object.</p> <p>Most influential factors: <i>Brand or manufacturer, reviews or word of mouth referrals, warranty or customer support</i></p>
Day 1 – Month 3	<p>Courting period: People project their existing mental models of trust from past experiences with similar objects from that brand or category onto the new object. Initial distrust tends to be highest during this phase when people have developed distrust toward similar objects, given the lack of contrary evidence that the new object addresses trust issues. For trusted objects, initial trust tends to be the highest due to novelty effects, which can insulate trust from damage due to poor performance or user experience design early on.</p> <p>Most influential factor: <i>Simplicity or transparency</i></p>
Month 3 – 6	<p>Make it or break it phase: As the novelty effect wears off, recurring issues with consistency, unreliability, or durability which may have been overlooked early on now become more prominent and noticeable. Trust/distrust are more vulnerable to change during this time. For frequently-used objects, negative perceived characteristics can result in a rapid breakdown of trust. During these months, people put the object through incremental “tests,” assessing how well the object performs at the jobs it was “hired” to do [4]. This is where learned trust [10] develops. As issues arise, people are willing to invest resources into resolving them if the perceived benefits are sufficient. Otherwise, particularly for low-cost or easily replaceable objects, trust can be irreversibly damaged, potentially resulting in disposing or discontinuing use of the object.</p> <p>Most influential factor: <i>Product performance, privacy and security</i></p>
Month 6 – 12	<p>Economic relationship: As the relationship nears the one-year mark, an economic model characterizes people’s trust/distrust calculus. If the object provides adequate convenience, improvement to everyday life, fun, or other valuable functionality, minor issues can be overlooked or forgiven. People may discover new features or functionality in addition to their core use cases which further builds trust. For trusted objects, trust level tends to stabilize at this point. For distrusted objects, people may adapt workarounds to deal with shortcomings if they perceive the net benefits to be positive.</p> <p>Most influential factor: <i>Product quality</i></p>
Year 1+	<p>Interdependence: For distrusted objects, relationships typically only remain only if they are attached to another object that cannot be replaced or there is another significant barrier to replacing or disposing of the object. In other words, people may “settle for less” and use the object in spite of their distrust. For trusted technology-related objects, many feel performance decline is a natural and reasonable expectation with age. Some will even commit to replacing it with the same object or brand in the future because it has “earned” their long-term trust and loyalty. Thus, for some object types, trust can be insulated from damage due to performance issues in the later stages of the relationship. In some cases, trust has an emotional dimension and anthropomorphic qualities may be attributed to the object.</p> <p>Most influential factor: <i>Time and personal experience</i></p>

factors were most salient and important to trust/distrust development. Table 1 presents the emergent themes of our analysis. Our results suggest a non-linear relationship between trust and time. We also identified the 3–6 month period of the relationship to be most vulnerable to fluctuation in people’s overall levels of trust/distrust.

In the first month of the relationship, many participants reported their trust was most resilient to stressors (e.g., privacy concerns, usability issues, poor performance or accuracy). However, following month one, participants reported their trust was much more vulnerable to damage:

“The size [of the device] was not even noticed at the beginning. I’m not sure why. It just didn’t bother me and I really enjoyed the device. **But then after around the three, maybe four, month mark, I think that’s when the honeymoon phase kind of tapered off.** I started to get really frustrated with some of the issues. I started noticing the battery life was not as good as it was at the beginning. Finally, the last straw was the size. When it really started to bug me, when it started to physically hurt my hand to stretch constantly across the screen, I was like, OK, that’s it.” –Man, 27 years old, discussing distrusted smartphone

Between 3–6 months, the trust relationship tends to be vulnerable. A frequently discussed issue influencing trust/distrust in agents (or smart objects) during the 3–6 month period was false positives. In particular, when these errors were not easily understood or explainable, this brought up significant concerns and scrutiny, even contributing to the final breaking point in distrustful relationships:

“I found some convenience in it and I really liked [the smart speaker] for a while ... And then I noticed, more often than not, it would cue or turn on, not because of its command word, but it just would. And it’s not even that its command word was said on TV or anything like that. No, it would just ‘wo-oh’ (*mimicking the sound of the device being activated*), and it’s just on, which made me very distrustful. **If it does that when we’re not even saying the command word, how do I know that it’s not listening all the time? And I just I don’t like that.**” –Man, 40 years old, discussing distrusted smart speaker

Months 6–12 were best characterized by an economic model of trust. Participants continued to develop experience with the object to understand whether the resources (e.g., their time, effort, money, personal data) they invested into the relationship were worth the perceived benefits. Evaluation of trust/distrust based on this cost-benefit analysis was the most frequently-reported strategy during this stage. Participants described imperfect expectations for technology-related objects, explaining they were willing to invest some time and effort into resolving issues. As long as they perceived a return on investment, trust was not necessarily damaged:

“Sometimes, it’s something like with the [smart] watch, [where] it’s a home run right off the bat. Everything that I expected it to do, it does. There might be a couple of hiccups here and there, but for the most part it needs and exceeds all of my needs. But then

there are other things that, you know, **sometimes there's a little bit of a learning process. I kind of have to find a workaround for things, or I have to do a little bit more research. It doesn't always end in distrust.** Sometimes it's a learning process, it kind of just depends on what it is. Some are a success right away, some are a gradual success, some are not successes at all." –Man, 42, discussing trusted smart watch

When the perceived costs were determined to be more significant than the perceived benefits, participants described an irreparable breakdown of their trust. This developed distrust sometimes resulted in object disuse or disposal:

"I had to do a bunch of research on forums to see what was going on, reach out to tech support, and it still didn't answer my questions or help me with what's going on. So, that's really why our relationship had to end, because it's just not reliable." –Man, 47, discussing distrusted smart home automation system

Finally, the most stability in people's trust/distrust levels were observed for relationships lasting one year or more. By this point in time, people gained sufficient familiarity and experience to understand the net benefits. For distrusted objects with high barriers for replacement, people developed strategies for relationship maintenance in spite of persistent negative issues.

"If something goes wrong, I always try to go further to try to solve it because I like to keep my products for as long as possible. And [technology company] was just, like, there is no solution... [A battery issue] started happening more frequently, and I would look up how to solve it and there were all these different solutions, and some of them worked, some of them didn't. By that two year mark I was like, 'I'm going to keep working with it' because I'm a college student, and I don't really want to buy a new laptop, and it still functions if I, like, just be very careful with it." –Woman, 24, discussing distrusted laptop

When objects performed reliably well over the course of several years, people reported a cyclical pattern between increased use, reliance, and trust. This long-term positive evaluation of the object's characteristics resulted in an extremely strong trust that was resilient to potential future decline, sometimes resulting in an emotional bond with the object. Some participants described extending this trust to the object brand or manufacturer as a result of the trustful long-term relationship, positively influencing initial trust toward new objects created by that entity.

"Dear [technology company] laptop, when I first opened you up it was like magic. The screen was beautiful, you worked super fast, and allowed me to accomplish all the tasks I needed for work over time. You have proven to be resilient by getting me through six years of professional work. As software updates have changed how you work and look, you still function the same way as day one. We have been through a

lot of changes over the years, with some bumps along the way. Some USB ports are not working as they once did, and you have slowed down a bit. But, you still continue to show me lots of lots of love by working well enough for me to continue to work. The fact that you're still working this well after such high usage over the years has made me trust you even more, to the point that I never want to leave you. I hope that this relationship and trust will continue to grow as we continue working together." –Man, 30, discussing trusted laptop

4.3 Mixed Methods Study

A dataset containing structured information about 420 agents was collected through a survey, of which 210 (50%) were classified as trusted and 210 as distrusted (50%). All participants were required to log information about at least 2 (1 trusted and 1 distrusted) agents, and a subset of participants entered additional agents, accounting for the additional agents in the dataset. For each agent, participants were asked to rank the top three most important reasons for their trust/distrust from a list of nine options. The nine options included the eight factors identified through analyzing the open-ended questionnaire data and one "other" write-in option. Analysis of our survey data revealed a ranking of the relative importance of these factors to building trust/distrust in agents (or smart objects). The most important reasons trust had been earned in these relationships (from most to least frequently selected by participants) are:

- (1) **Product performance (accuracy, precision, reliability)** (67.6%)
- (2) **Time and personal experience** (56.7%)
- (3) **Simplicity or transparency (understanding how it works)** (48.1%)
- (4) Brand or manufacturer (41.0%)
- (5) Product quality (materials, durability, aesthetics) (38.6%)
- (6) Security and privacy (30.5%)
- (7) Reviews or word of mouth referrals (11.9%)
- (8) Warranty or customer support (4.3%)
- (9) Other (1.4%)¹

The most important reasons leading to distrust (from most to least frequently selected by participants) are:

- (1) **Security and privacy** (59.0%)
- (2) **Product performance (accuracy, precision, reliability)** (56.7%)
- (3) **Time and personal experience** (52.4%)
- (4) **Simplicity or transparency (understanding how it works)** (37.1%)
- (5) Product quality (materials, durability, aesthetics) (31.0%)
- (6) Brand or manufacturer (23.8%)
- (7) Reviews or word of mouth referrals (23.8%)
- (8) Other² (10.0%)
- (9) Warranty or customer support (6.2%)

¹We analyzed the "other" responses for the top reasons for trusting smart agent. 2 out of the 3 responses were related to "Security and privacy."

²We analyzed the "other" responses for the top reasons for distrusting a smart agent. 13 of the 21 responses were related to "Security and privacy."

We conducted a Pearson's Chi-squared test to determine if the distribution of the most mentioned trust factors is different from the distribution of the most mentioned distrust factors, $X^2 = 58.3$, $df = 8$, $p < .001$. The results indicate people assign a different importance to factors with regards to the development of trust (with product performance being the top factor) compared to distrust (with security and privacy being the top factor) in their relationships with agents (or smart objects). This suggests HAI designers need to treat building trust and preventing distrust as two different goals, attending to different factors contributing to each.

5 DISCUSSION AND CONCLUSION

This research makes novel methodological and empirical contributions to the study of trust in HAI. Our studies examined human-object relationships, yielding insights on trust/distrust development from a broader superset of relationships that included but were not limited to human-agent trust. Based on the rich qualitative insights on long-term trust development we gained using this approach, we propose future work further investigate how empirical research and frameworks on human embodied interaction with everyday objects can be applied to HAI research. This broadened lens becomes more relevant as the boundaries between "intelligent" and "non-intelligent" objects are increasingly blurred by emerging technologies such as ubiquitous computing, the internet of things, augmented reality, and wearable computing.

Our findings suggest human-object and human-agent trust/distrust development are based on a continuous evaluation of characteristics over time and are influenced by the perceived resource exchange between the human trustor and the trustee (i.e., the object, technology, smart device). The salience and importance of different internal and external factors influencing trust were found to vary based on the time stage of the relationship. We found trust-related characteristics of an agent are perceived differently based on historical experiences of how well they met (or failed to meet) people's changing expectations over time. We recommend a primary objective of HAI designers is to establish and maintain an upward trajectory of trust in the long run, rather than trying to maximize trust level at the onset of the relationship. From an ethical standpoint, it is critical for designers to address people's needs and concerns while keeping them well-informed throughout the trust building process.

While first impressions were important to the trust-building process, it was even more critical to continue to develop trust after novelty effects expired. Our results suggested trust levels were most vulnerable and volatile during the 3–6 month period. HAI designers must ensure the object's desired characteristics can be sustained throughout regular interactions, with quality control standards being met for a minimum of six months. In spite of the relationship's vulnerability to trust breakdown during this period, following it, people were willing to invest resources into resolving issues if the perceived benefits outweighed the costs. We recommend the development of strategies to monitor the net exchange of resources to and from the human trustor and to ensure the perceived net benefits remain positive at the risk of damaging trust. When people invest personal data resources, designers should use strategies to proactively communicate the potential benefits and address perceived risks throughout these transactions. Future work should

examine trust relationships between the 3–6 month time frame in more detail to understand what causes this vulnerability and identify potential signals which indicate trust decay to aid in generating novel design strategies for trust repair.

We found brand or manufacturer and warranty or customer support were significant factors to the development of trust/distrust. In spite of these results lacking epistemological attention in traditional academic literature on HAI, we argue these factors should not be taken for granted by designers. Participants frequently personalized brands and manufacturers in the study, referring to the names of the technology companies who manufactured their agents (rather than the names of the agents themselves) when discussing trust/distrust. Despite brand image and customer support not being innate attributes of technology design, these factors play a critical role in how they perceived agent trustworthiness. Designers cannot ignore their importance, particularly in applied contexts such as industry settings.

Finally, we identified product performance to be the most important trust-building factor and security and privacy to be the most important distrust-building factors. Trustworthy agents must tread a careful line ensuring privacy and security concerns are addressed while providing them with a highly accurate, precise, and reliable experience. For systems relying on personal data to improve accuracy, this is an especially challenging balance to strike. This finding aligns with prior work on the *Personalization Privacy Paradox* [1], which describes the tension between needing personal data to improve performance and personalization while addressing people's concerns about privacy and data security. Rather than developing strategies to promote data sharing despite privacy concerns, we argue that trust should be earned by directly addressing root causes of these concerns and remaining transparent throughout the relationship. Designers must also evaluate how different privacy settings will affect overall performance. This information should be disclosed in a proactive and straightforward way to inform and support user choice. Trust attitudes and situational factors greatly influence people's privacy calculus [11], and a better understanding of how imbalances between trust in the agent, desired level of personalization, and personal data disclosure behaviors interact and affect long-term relationship management is needed. Future work should more closely examine these factors between humans and agents in people's natural settings over long periods of time.

Like other types of partnerships people have, the human-agent partnership is dynamic and constantly evolving. During different time periods, people weigh influential factors differently to evaluate their trust/distrust, and they adopt resourceful strategies to try to maintain trust. We hope that our evidence-based implications and recommendations for HAI designers and practitioners will aid in the research and development of agents which foster trustful long-term relationships.

ACKNOWLEDGMENTS

We thank the following individuals for their contributions to this research: Dr. Mengjie Yu, Dr. Meiyuzi Gao, Vivien Francis, and Dr. Simon Wolfe Taylor. We also thank the anonymous reviewers for their feedback.

REFERENCES

- [1] N. F. Awad and M. S. Krishnan. 2006. The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to be Profiled Online for Personalization. *MIS Q.* 30 (2006), 13–28.
- [2] Rachel K. E. Bellamy, Sean Andrist, Timothy Bickmore, Elizabeth F. Churchill, and Thomas Erickson. 2017. Human-Agent Collaboration: Can an Agent Be a Partner?. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI EA '17). Association for Computing Machinery, New York, NY, USA, 1289–1294. <https://doi.org/10.1145/3027063.3051138>
- [3] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. sage.
- [4] Clayton Christensen and Michael Raynor. 2013. *The innovator's solution: Creating and sustaining successful growth*. Harvard Business Review Press.
- [5] Mary Czerwinski, Eric Horvitz, and Susan Wilhite. 2004. A diary study of task switching and interruptions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 175–182.
- [6] dscout Inc. 2020. *In-context insights via remote qualitative research*. <http://www.dscout.com>
- [7] Thomas Erickson. 1997. *Designing Agents as If People Mattered*. MIT Press, Cambridge, MA, USA, 79–96.
- [8] Umer Farooq and Jonathan T. Grudin. 2017. Paradigm Shift from Human Computer Interaction to Integration. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI EA '17). Association for Computing Machinery, New York, NY, USA, 1360–1363. <https://doi.org/10.1145/3027063.3049285>
- [9] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in personality* 37, 6 (2003), 504–528.
- [10] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* 57, 3 (2015), 407–434.
- [11] Flavius Kehr, Tobias Kowatsch, Daniel Wentzel, and Elgar Fleisch. 2015. Blissfully ignorant: the effects of general privacy concerns, general institutional trust, and affect in the privacy calculus. *Information Systems Journal* 25, 6 (2015), 607–635.
- [12] John Lee and Neville Moray. 1992. Trust, control strategies and allocation of function in human-machine systems. *Ergonomics* 35, 10 (1992), 1243–1270.
- [13] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, Vol. 53. Citeseer, 6–8.
- [14] Michael Wooldridge and Nicholas R. Jennings. 1995. Intelligent agents: theory and practice. *The Knowledge Engineering Review* 10, 2 (1995), 115–152. <https://doi.org/10.1017/S0269888900008122>
- [15] Naim Zierau, Christian Engel, Matthias Söllner, and Jan Marco Leimeister. 2020. Trust in Smart Personal Assistants: A Systematic Literature Review and Development of a Research Agenda. (2020).