The Effects of Past Experience on Trust in Repeated Human-Agent Teamwork

Socially Interactive Agents Track

Feyza Merve Hafızoğlu
Üsküdar University
İstanbul, Turkey
feyza.hafizoglu@uskudar.edu.tr

Sandip Sen
The University of Tulsa
Tulsa, OK
sandip@utulsa.edu

ABSTRACT

For human-agent virtual ad hoc teams to be effective, humans must be able to trust their agent counterparts. To earn the human's trust, agents need to quickly develop an understanding of the expectation of human team members and adapt accordingly. This study empirically investigates the impact of past experience on human trust in and behavior towards agent teammates. To do so, we developed a repeated team coordination game, the Game of Trust (GoT), in which two players repeatedly cooperate to complete team tasks without prior assignment of subtasks. The effects of past experience on human trust are evaluated by performing an extensive set of controlled experiments with participants recruited from Amazon Mechanical Turk, a crowdsourcing marketplace. We collect both teamwork performance data as well as surveys to gauge participants' trust in their agent teammates. The results show that positive (negative) past experience increases (decreases) human trust in agent teammates and past experience can affect three antecedents of trust: emotional state, game expertise, and expectation. These findings provide clear and significant evidence of the influence of key factors on human trust in virtual agent teammates and enhance our understanding of the changes in human trust in peer-level agent teammates with respect to past experience.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**;

KEYWORDS

Human-Agent teamwork; trust; past experience; virtual environment

ACM Reference Format:
Feyza Merve Hafızoğlu and Sandip Sen. 2018. The Effects of Past Experience on Trust in Repeated Human-Agent Teamwork. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 9 pages.

1 INTRODUCTION

As agent capabilities have improved along several key functionalities, such as the ability to plan, collaborate, and coordinate, agents are being increasingly accepted as *partners* in collaboration with humans. Engaging in group activities with intelligent virtual agents (hereafter referred to as "agents") have become prevalent in various domains. Enabling, supporting, and improving human-agent interactions, particularly virtual ones and those over large distances, will then offer notable, practical benefits to individuals and to the society as a whole in diverse scenarios.

Trust as social glue plays a critical role in inducing cooperative behavior among individuals and within groups [3, 25, 29, 30]. The role of trust is not limited to human interactions: trust also shapes the way people engage with technology [8, 18, 25, 39]. Therefore, establishing people's trust is a key cornerstone of fluid interactions between humans and agents. In particular, we need to develop agent technology that will enable developing agent applications in domains where humans recognize agents as *autonomous* and *effective* partners and have to rely on agents as "peer" level team members.

This study aims to better understand how human trust in agent teammates changes over repeated interactions in the context of virtual human-agent teamwork. By *virtual* human-agent teamwork we refer to domains where autonomous agents and humans work over a network without any physical embodiment of the agents, either in the form of robots or avatars. We consider human trust behavior which is based only on the agent's task performance or contribution towards achieving team goals over repeated interactions.

In such virtual human-agent teamwork domains, human trust attitudes will be influenced by a variety of factors, including agent reliability, prior experience(s) of humans, and agent reputation. In this study, we explore *past experience* of humans, which refers to prior interactions with other agents within the same or similar domains. People use their knowledge from past interactions with others for assessing the trustworthiness of a trustee, such as a person, an organization, or a computer agent [6, 10, 28]. Thus an individual's past experience can affect the perceived trustworthiness of agent teammates and the outcome of the teamwork [16].

The central question of this study is: How do repeated interactions with a given agent teammate alter the initial trust development of an individual when interacting with a subsequent agent teammate? What kind of past experience increases an individual's inclination to trust an agent teammate? What are the effects of experience with an untrustworthy agent on the subsequent interactions with other agent teammates?

We investigate these issues to better understand how to augment agents with the necessary capabilities so that they can effectively collaborate with human teammates. We developed a virtual teamwork game where participants interact for a small number of teamwork situations with an agent. In each interaction, the participant knows about the total work units to be performed to achieve the

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

team goal and has to choose its effort without explicitly coordinating with its teammate. The work effort of the teammate and the combined team performance are revealed to the players after the game. We collected data on work effort choices and team performance and also surveyed the participants' trust perception of their agent teammate. The analysis of this data enables us to infer the effects of work efforts by the agent teammate on the participants' trust and on the resultant choice of work effort by the participants. We performed experiments with the participants where they were involved in several games with different agent teammates. The goal of these experiments is to observe how past experience biases a participant's trust in their teammates in subsequent games.

The rest of the paper is structured as follows. Section 2 presents the related work. Section 3 describes the human-agent teamwork model that is considered in this research, while Section 4 explains our empirical methodology. In Section 5, we present the results of experiments and discuss the empirical findings in Section 6. Finally, Section 7 concludes the paper with a summary and the directions for future research.

2 RELATED WORK

The importance of trust in human-agent interactions has been well-acknowledged in literature [12, 17, 20, 23, 40]. The factors affecting human trust in agents can be grouped into three broad classes: human factors (as trustors), agent factors (as trustees), and external factors (environment). Various studies have investigated the effects of human factors, such as age [6, 35], personality [13], culture [21], mood [38], attitude [31], and past experience [6, 10, 28].

Previous research demonstrates that the effect of past experience on human trust behavior towards technology differs between context. Manzey et al. [28] discover that negative past experience causes operators to have reduced trust in automation, while Chen et al. [10] found that past experience with websites enhanced the perceived ability, integrity, benevolence, and predictability of e-vendors. Dutton et al. [15] suggest that positive (negative) past experiences led to trust more (less), however, this influence diminishes over time. There is a relatively small body of literature that is concerned with the past experience factor in the context of human-agent teamwork.

Agent behavior is fundamental in building trust in agent teammates. Positive behavior, such as cooperativeness [41] and reliability [17], improves trust and facilitates the collaboration between parties. In contrast, negative behavior, such as defection [40] and deception [41], leads to reduced trust and, hence, less willingness to collaborate in future interactions. Communication skills of agents play a significant role in maintaining the trust relationship [20, 33, 42]. Furthermore, familiarity and personalization of agents have been shown to positively influence human trust [27, 43].

In addition to agent behavior, researchers have investigated the effects of different agent representations, such as avatars and robots [1, 11, 36, 43], and the effects of external factors, such as information representation [4] and reputation [19].

The vast majority of studies on human-agent teamwork assumes that team members can coordinate their actions either through communication or pre-defined protocols, such as commitment [22], negotiation [40], giving advice [12, 38], providing recommendations [27], and physical interaction [36].

Recently, new environments, that enable group activities or collaboration between humans and agents, have been emerging, such as crowd-work with complex tasks [26] and massively multiplayer online games [9]. In such environments, humans collaborate with peer level agent teammates to achieve a common goal without pre-planning. This kind of human-agent teamwork, without explicit prior coordination, has been rarely investigated from the angle of human trust. In a study on human-agent teamwork without explicit coordination, Merritt et al. [32] examined the blame behavior for team failures. In another study, Ong et al. [34] demonstrate that a cooperative representation of the game improves trust in agent teammates compared to a competitive representation.

Our research extends these studies on human trust in technology as follows: considering teams of human and agent rather than mere interactions between two players [1, 11, 41, 43]; focusing on teamwork environments in which there is neither explicit communication between human and agent (as in [22, 40]) nor agents embodied in physical forms, such as robots (as in [1, 11, 22, 43]); exploring repeated, in contrast to one-shot [42], interactions in fixed rather than dynamic teams [40]; providing real team tasks for evaluating human-agent teamwork rather than the standard artificial environments [1, 11, 22, 40–42]. To the best of our knowledge, this is the first study on past experience affecting trust in human-agent teamwork without prior coordination within a repeated virtual team game scenario where agents are peer-level teammates.

3 HUMAN-AGENT TEAMWORK MODEL

Our goal is to understand and characterize human trust development in agent teammates over initial repeated interactions, but without any prior experience of interaction with that agent, in the following scenarios:

- The individual is new to a domain and has to rely on more experienced agent teammates until she develops the necessary competency from her own experiences,
- The individual is familiar with the domain but will need to work with autonomous teammates, with whom the individual has had no prior collaboration experience, to be able to process task assignments beyond their own capacity.

In such domains including ad-hoc teamwork scenarios, unfamiliar individuals have to cooperate with new partners. Such cooperation can be engendered by time-critical responses to emergency situations, as well as by the need to find effective partners to complement the capabilities of dynamically changing teams, e.g., humans or agents leaving the system or switching to other groups. In a number of such scenarios, the capabilities and trustworthiness of new partners for contributing to team goals are at best partially known. Additionally, extensive pre-planning may not be possible to optimally allocate dynamically arriving tasks among team members. Rather, the team must be responsive to the emerging situations that can be achieved by team members adapting their behaviors and efforts based on expectations of contribution by team members.

In this context, we use the following operational characterization that captures what it means for a human to trust an agent teammate: *Trust in an agent teammate reduces the uncertainty over that agent's* *independent actions which positively correlates with the truster's utility towards achieving team goals* [37]. Based on this interpretation, human trust in an agent teammate can both reduce uncertainty about agent's contribution and improve team performance through more efficient team coordination.

3.1 The Game of Trust

The Game of Trust (GoT) is a two-player team game where each pair of players partake in n sequential interactions. In the i^{th} interaction, players are assigned a team task, t_i. The team task consists of $|t_i|$ atomic subtasks of the same type, hence $|t_i|$ is the size of the team task. There are no dependencies between the subtasks. We assume these subtasks do not require any specialized skills and hence both the human and the automated player can accomplish them if they wanted to. Examples of such tasks with undifferentiated subtasks, where only the number of subtasks accomplished by the team matter, include recruiting a given number of volunteers, collecting a number of specimens that fit a given description, and so on.

There is no prior assignment of subtasks to players nor are the players allowed to communicate to select subtasks. Instead, each player decides how many subtasks she will perform individually given the size of the team task, $|t_i|$, without knowing the number of subtasks that the other player will perform. After separately performing subtasks, players are told whether the team has achieved the team goal, i.e., whether the two players combined have completed the required number of subtasks, as well as the number of subtasks that the other player completed.

There is a cost of performing subtasks that is computed by the cost function, c, based on the number of subtasks completed. Both players have their individual payment accounts, from which they can pay for the cost of performing tasks, which have an initial balance of b_{init} at the beginning of the game. Players are instructed about the cost and reward functions. The cost of the subtasks that are performed by each player is withdrawn from the corresponding account. If the combined number of subtasks accomplished by the players is equal to or greater than the size of the team task, it means that the players successfully completed the team task. In that case, the reward computed by the reward function r is equally split between players and deposited to their individual accounts. If, however, the combined number of subtasks that the players accomplished is less than the team task, no reward is given.

By *utility of a player* we refer to half of the team reward, if any, minus the cost of performing subtasks individually. If they cannot achieve the team task, both players may lose utility from this teamwork instance. Even if they achieved the team task, a player loses utility if the cost of the player's performance is greater than half of the team reward. Finally, *social utility* corresponds to the sum of the utilities of the two players. Social utility is optimized when the total number of subtasks completed by team members is precisely equal to the team task size.

3.2 Domain Description

In our study, a team consists of one human and one agent playing the *Game of Trust*. We did not want team task to require any specialized skills that may impose extra constraints and undue burden on participants. Furthermore, our goal was to choose task types that are neither particularly boring nor particularly attractive[1]. Based on these considerations, we chose an audio transcription domain for the human-agent teamwork goal instances. In this domain, the *task* that is assigned to the team corresponds to transcribing a number of words and the *atomic subtask* corresponds to transcribing one word.We will use the term *task size* to refer to the number of words to transcribe, i.e., number of subtasks, in an interaction.

The purpose of the transcription task is to mimic a real teamwork environment where the participants have to collaborate with their automated teammate to achieve their shared goal which they cannot achieve by themselves. Though we have no interest in the transcribed words, the participants are still required to transcribe a word with at least 60% accuracy to receive credit for successful transcription. We compute the dissimilarity between the transcription and the transcribed word as the edit distance[2] over the length of the transcribed word. This is done to ensure a minimum quality of participant effort. Inaccurate transcriptions are not counted but their cost is withdrawn from the player's budget.

We require one human player to play a series of games, where each game consists of a sequence of interactions with one of several automated player types. Both human and agent players are expected to be self-interested: the more words a player transcribes, the higher the player's cost is. Subsequently, higher cost leads to a lower player utility. On the other hand, the less they perform, the higher is the risk of not achieving the team goal. Therefore, the number of words they need to transcribe is a critical decision that they have to make in each interaction and is based on their trust in the teammate for contributing to the team task.

4 EMPIRICAL METHODOLOGY

Past experiences are grouped into two broad types: *positive* and *negative*. By positive (negative) experience, we refer to engaging with trustworthy (untrustworthy) agent teammates in past teamwork instances. We expect that interacting with a (an) trustworthy (untrustworthy) agent teammate is most likely to inspire positive (negative) feelings. It is shown that happiness significantly improves trust whereas anger significantly reduces it [14]. This finding is further supported by previous studies on the relation between the essence of the experience, and resulting trust behavior [10, 28]. Therefore, we posit that positive (negative) past experience leads participants to perceive their agent teammate more (less) trustworthy:

Hypothesis 1. Positive past experience increases initial trust in future agent teammates.

Hypothesis 2. Negative past experience decreases initial trust in future agent teammates.

Besides trust, the relationship between past experience and human behavior, effort levels, is of our interest as well. We argue that the influence of negative experience on effort levels may be twofold. First, negative feelings based on negative experience may reduce participant's enthusiasm to expend effort, whereby participant may tend to deliver less work. Second, negative experience may lead

[1] This facet was considered to avoid, to the extent feasible, the possibility of participants having additional motivations that either positively or negatively biased their choice of effort level or contribution to the team goal.
[2] http://en.wikipedia.org/wiki/Wagner-Fischer_algorithm

participants to play more cautiously and deliver more work. These two effects engender biases that are at odds with each other. On the other hand, positive experience may facilitate relying on agent teammate, i.e., being less cautious, for achieving the team goal, hence deliver less work. Based on these divergent possibilities, we expect past experience to have a tangible influence on participants' effort levels in subsequent team participation.

Hypothesis 3. Past experience with an agent teammate affects effort levels by the participants in subsequent interactions with other agent teammates.

4.1 Agent Teammates

In order to realize positive and negative experiences, we developed two agent players that resemble trustworthy and untrustworthy behavior. Playing the GoT with the trustworthy (untrustworthy) agent corresponds to positive (negative) experience in this study.

4.1.1 Trustworthy Agent Player. Given that teammates delivering half or more of the team task are perceived to be fair and trustworthy, the *trustworthy player* initially delivers half of the team task and thereafter increases its effort level if the previous interaction was a failure. Formally, the number of work units completed by the trustworthy agent in i^{th} interaction is

$$w^i_{Trustworthy} = \frac{t_i}{2} + \Delta^i,$$

$$\Delta^i = \begin{cases} 0 & \text{if i=1} \\ \Delta^{i-1} + 1 & \text{if } w^{i-1}_h + w^{i-1}_{Trustworthy} < t_{i-1} \\ \Delta^{i-1} & \text{otherwise.} \end{cases}$$

where t_i is the team task size of i^{th} interaction, w^{i-1}_h is the number of subtasks complete by the human in the $(i-1)^{th}$ interaction, and Δ^i (initially zero, i.e., $\Delta^1 = 0$) is the surplus work to fair share in i^{th} interaction.

4.1.2 Untrustworthy Agent Player. We designed an *untrustworthy player* that is neither a dummy agent, e.g., randomly making unfair choices, nor a smart exploiter, e.g., optimizing the social utility by completing just the necessary amount of work. Our intention is to ensure participants believe their teammate is inclined to exploit them whenever they have a chance, e.g., reducing its efforts when human consistently delivers more than a fair share. The untrustworthy player makes at least one unfair choice in a game. The number of work units delivered by the untrustworthy player in i^{th} interaction is

$$w^i_{Untrustworthy} = \frac{t_i}{2} - \Delta^i.$$

The amount of deviation from the fair share in ith interaction, Δ^i, is stochastically incremented. Therefore, its effort is monotonically non-increasing[3] and decreases occasionally. Algorithm 1 describes the task size choice function of the untrustworthy player.

[3]There are two exceptions to this facet of the untrustworthy player: (1) if the team failed in the last three interactions, the untrustworthy player completes half of the team task, and (2) if the team failed in the last two interactions, the untrustworthy player delivers half of the team task or half of the team task minus one.

Algorithm 1: Task size function of the *Untrustworthy Agent*

Input : t_i, team task size;
nFailures, number of failures in the game;
Δ, a global variable initialized to 0 in the game;
p_{min}, a global variable, to set the minimum value of the parameter p, initialized to 0.25 in the game;
Output: $w^i_{Untrustworthy}$, the task size choice

1 **if** $nfailures \geq 3$ **then**
2 $\quad \Delta \leftarrow 0$
3 **else if** $nfailures \geq 2$ **then**
4 $\quad \Delta \leftarrow x$ `// random number x ∈ [0,1]`
5 **else if** $i > 3$ *and* $\Delta = 0$ **then**
6 $\quad \Delta \leftarrow 1$
7 **else**
8 $\quad p \leftarrow p_{min}$
9 $\quad \epsilon \leftarrow 0$
10 $\quad$ **if** $i > 1$ **then**
11 $\quad\quad \epsilon \leftarrow \frac{w^{i-1}_h}{t_{i-1}} - 0.5$
12 $\quad$ **if** $\epsilon > 0$ **then**
13 $\quad\quad p \leftarrow p + \epsilon$ `/* Increase the probability to increase Δ */`
14 $\quad$ **if** $rand(0, 1) < p$ **then**
15 $\quad\quad \Delta \leftarrow \Delta + 1$
16 $\quad\quad p_{min} \leftarrow p_{min} - 0.05$ `/* Higher the value of Δ, lower the probability to increment Δ */`
17 $w^i_{Untrustworthy} \leftarrow \frac{t_i}{2} - \Delta$
18 **return** $w^i_{Untrustworthy}$

The first two conditions prevent being perceived as an imprudent player. When the team experiences a number of recent failures, a reasonable player's reaction is to increase its effort. To do so, the untrustworthy agent completes half of the team task, i.e., $t_i/2$, if the recent three interactions were failures. Likewise, it completes half of the team task or half of the team task minus one if the last two interactions were failures.

The third condition (line 5-6) ensures that the untrustworthy agent exhibits untrustworthy behavior at least once. If the value of Δ has not been incremented so far, i.e., the untrustworthy agent has delivered half of the team task, it will deliver less than the fair share by incrementing the value of Δ.

In the else condition, *Bernoulli distribution* is used to determine whether the value of Δ will be incremented (line 14). In an interaction, the base value of the parameter p is initialized with p_{min}, a global variable (line 8). If the participant delivers more than the fair share in the previous interaction, the value of p is increased by the value of excess effort of the teammate. That means, the higher the effort level by the teammate, the higher is the probability to increase the value of Δ, i.e., delivering less work. In order to prevent even higher values of Δ, the value of minimum probability to increment Δ, p_{min}, is subsequently reduced by 0.05 (line 16). Finally, the individual task size is computed as half of the team task minus Δ (line 17).

4.1.3 Learner Agent. *Learner agent* is trained offline to predict human player's task choices by utilizing the linear regression with the data collected from the teamwork experiences of humans from previous experimentation. It delivers half of the team task size in the first interaction. Subsequently, given an accurate prediction of teammate's task choice based on prior interactions with the human player, the learner agent chooses to complete the rest of the team task to achieve the team goal optimally and without redundancy or falling short of the team goal.

4.2 Experimental Setup

Game Configuration: The number of interactions in a game is five (as in [5, 7, 34]), which is short enough to avoid participants becoming bored while allowing team members to adapt to teammates with predictable behavior. The size of team task is incremented by two in each interaction, i.e., the sequence of task sizes is $\langle 6, 8, 10, 12, 14 \rangle$.

Both the participant and the agent have their private account with an initial balance of 45, which is sufficient to complete all the tasks in the sequence. The cost and reward per work unit are set to 1 and 1.75, respectively. The players are allowed to choose a task size between one and the size of team task minus one.

Experimentation: Each experiment consists of two games, where the first game is for providing the user with past experience and the second game is for investigating the effects of that experience on subsequent interactions with another agent. We experimented with two groups of participants based on their experience, *G1* and *G2*, with the associated teammate orderings:

G1: Trustworthy Agent, Learner Agent;
G2: Untrustworthy Agent, Learner Agent.

At the time of playing the second game, playing with the trustworthy (untrustworthy) agent in the first game resembles having positive (negative) past experience, respectively, for the participants. In the second game, we paired the participants with the *Learner* agent, because we anticipated these interactions to be shaped by the participants' biases based on their past experience. Thus we eliminated other factors that are within our control of this study, such as the order of the game and the trustworthiness of the agent, as these factors may also affect the perceptions and decisions of the participants. We investigate the influence of prior experience by comparing the results from the second game played by the participants in *G1* and *G2*. Between the two groups, the only difference[4] was their past experience; the order of the game and the agent teammate were the same for both groups.

Survey: The game includes a short survey on trust to assess the participants' perceived trustworthiness and fairness of their teammates. Participants completed this survey after the first, third, and fifth interactions of a game (similar to [38]) after they were shown the outcome of the most recent teamwork. This short questionnaire, adapted from [2], consists of the following items which are rated on a 5-point Likert scale from "Strongly disagree" to "Strongly agree": (1) I trust my teammate and would like to continue to participate in other teamwork with my teammate, (2) My teammate is fair in performing team tasks, and (3) My teammate works responsibly for accomplishing the team task. The trust level of a participant in agent teammate is computed as the average of the responses.

[4]The participant population is different between the two groups, as it was not feasible to utilize the same population to test positive and negative experiences.

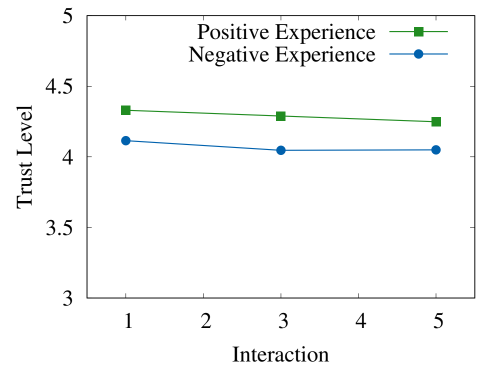


Figure 1: Trust in Learner Agent for positive and negative past experiences

Metrics: We adopted three essential metrics in our analysis. (1) In order to analyze the impact of past experience on participants' trust, we used *trust level* in an agent teammate that is computed as the average of the participants' average responses to the first three survey items. (2) To analyze how participants' behavior is affected, we used *effort level* that is the portion of the total work units completed by this team member, i.e., the fraction of individual task size over the team task size, and has a value in the range $[0, 1)$. (3) To analyze the relationship between past experience and team performance, we used *social utility* (see Section 3.1) as well as the cumulative outcomes in a game.

Participants: We recruited 216 participants through Amazon Mechanical Turk[5]. Data of 16 participants was eliminated due to insufficient attention. There were 98 and 102 participants in groups G1 and G2, respectively. Approximately 50% of the participants were female. Age distribution was as follows: 18-24 years, 18%; 25-34 years, 47%; 35-44 years, 17%; 45-54 years, 13%; 55-64 years, 4%; and 65 years or older, 1%. The distribution of education levels was as follows: high school degree, 15%; some college experience, 29%; associate's degree, 10%; bachelor's degree, 29%; and graduate degree, 16%; and PhD, 1%. The ethnicity distribution was as follows: White, 79%; Hispanic-Latino, 5%; African-American, 6%; Asian, 7%; and other ethnicities, 3%.

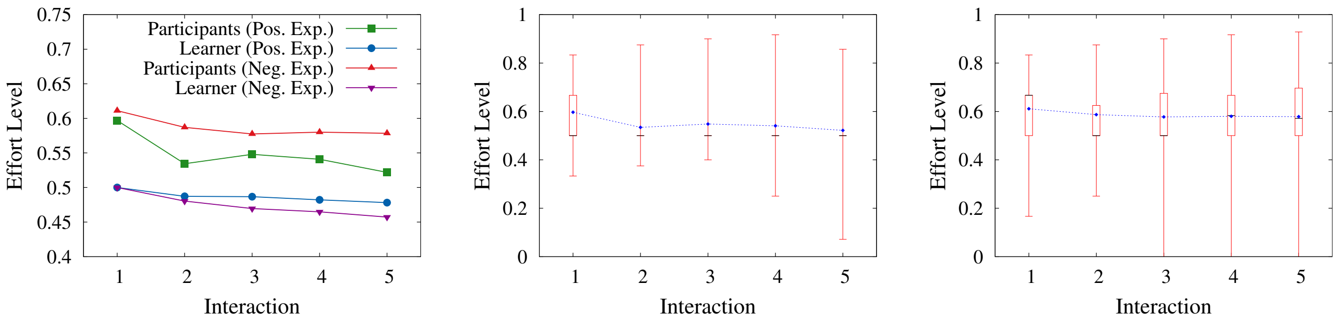
5 RESULTS

This section presents the *trust level*, *effort level*, and *team performance* analysis of experimental data that is collected from the second game played with the Learner agent. One-way ANOVA analysis is used to test the statistical significance.

5.1 Trust Analysis

Within Condition: Figure 1 depicts that the participants' trust in the Learner agent after first, third, and fifth interactions for the two

[5]http://www.mturk.com/



(a) Positive and Negative Experience (b) Positive Experience (c) Negative Experience

Figure 2: Effort levels by participants and Learner Agent and quartile distribution of effort levels by participants for positive and negative past experiences

conditions. Trust in the agent teammate did not vary significantly over interactions for both positive and negative past experiences.

Between Condition: The participants differentiated between the conditions with respect to their trust in the agent teammate. Higher levels of trust for positive past experience condition demonstrate that positive experience facilitated the trust formation process. In particular, initial trust is significantly greater ($F(1, 198) = 4.98,\ p < 0.05$) for positive past experience ($M = 4.33,\ SD = 0.71$) compared to negative past experience ($M = 4.11,\ SD = 0.65$). However, the difference between the two conditions gradually decreases towards the end of the game. It is likely that the influence of past experience diminishes over time or that it is dominated by present experience.

5.2 Effort Level Analysis

Figure 2 presents the effort level distribution and the quantile distribution of effort levels.

Within Condition: Figure 2(a) depicts the effort levels by the participants and the Learner agent for the two conditions. The participants having positive past experience significantly decreased their effort levels over interactions ($F(4, 485) = 6.74,\ p < 0.001$). In the second interaction, effort levels were sharply reduced from 0.60 to 0.53. It is likely that positive past experience encouraged the participants to rely on the agent teammate and reduce their effort. Additionally, effort level by the Learner agent declined over interactions as well ($F(4, 485) = 2.06,\ p < 0.1$). For negative past experience, the variation in effort levels by the participants is not significant, whereas effort level by the Learner agent significantly decreased over interactions ($F(4, 505) = 3.86,\ p < 0.01$). Since the participants could not rely on the Learner agent, they expended greater efforts which, in turn, led the Learner agent to reduce its effort.

Between Condition: In Figure 2(a), when comparing past experiences, average effort levels were lower for positive past experience compared to those for negative past experience over the course of the game. Surprisingly, initial effort levels by the participants were not significantly affected by past experience. In the second interaction, however, effort levels were significantly lower ($F(1, 198) = 10.86,\ p < 0.01$) for positive past experience ($M = 0.53,\ SD = 0.09$) compared to those for negative past experience ($M = 0.59,\ SD = 0.13$). Similarly, effort levels for positive past experience ($M_4 = 0.54,\ SD_4 = 0.11;\ M_5 = 0.52,\ SD_5 = 0.12$)[6] were significantly lower than those for negative past experience ($M_4 = 0.58,\ SD_4 = 0.17;\ M_5 = 0.58,\ SD_5 = 0.17$) in the fourth ($F(1, 198) = 3.90,\ p < 0.05$) and fifth ($F(1, 198) = 7.55,\ p < 0.01$) interactions.

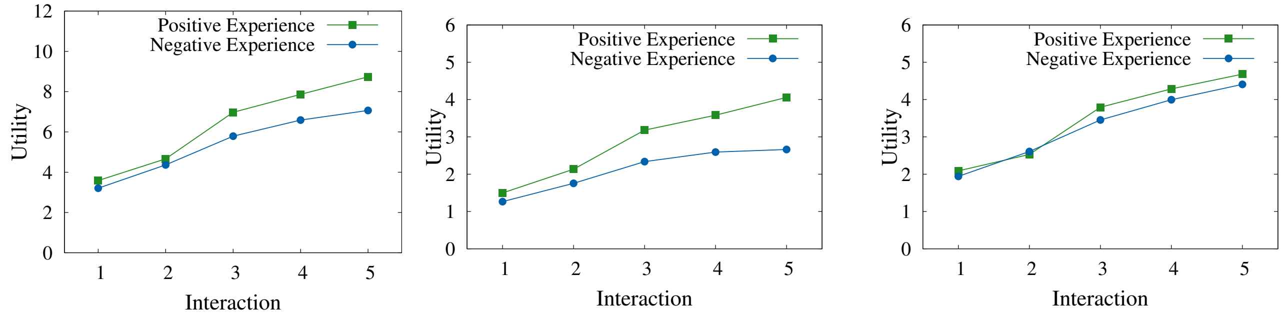
Overall the results indicate that past experience has an impact on the participants' effort levels as follows: negative past experience led the participants to make more cautious choices whereas positive experience encouraged more reliance on the agent teammate to complete the task and, thereby, reduced effort levels.

In Figures 2(b) and 2(c), we present the quartile distribution of effort levels by the participants for positive and negative past experiences. Figure 2(b) shows that positive past experience led the majority of the participants to deliver half of the work. In the first interaction, the upper quartile having a value of 0.67 illustrates that participants began playing cautiously despite their positive past experience. However, starting from the second interaction, the lower quartile, median, and upper quartile have a value of 0.50 for the rest of the game. This means that at least 50% of the participants delivered half of the team task in the last four interactions. The standard deviation of effort levels varies between 0.8 and 0.12 in the five interactions, which shows the homogeneity of the effort levels in the positive experience condition.

Higher effort levels of the participants engendered by negative past experience led to significantly greater redundancy compared to positive past experience ($F(1, 198) = 18.75,\ p < 0.001$). Subsequently, positive past experience led to significantly higher participant utility ($F(1, 198) = 11.85,\ p < 0.001$) and hence, social utility ($F(1, 198) = 7.07,\ p < 0.01$).

Figure 2(c) shows the quartile distribution of effort levels for negative past experience. Initially, at least 50% of the participants put in more than 0.67 effort. Initially, both the median and the upper quartile had a value of 0.67. Then the median dropped to 0.50 and the upper quartile dropped 0.63 in the second interaction. In the subsequent interactions, their values increased to 0.57 and 0.70,

[6] Subscripts after mean (M_i) and standard deviation (SD_i) report the interaction number i.



(a) Social utility (b) Participant utility (c) Agent utility

Figure 3: Utility distribution for positive and negative past experiences

Table 1: Results of the games for positive and negative past experiences

	Positive Past Experience	Negative Past Experience
Goals Achieved	4.76 ± 0.62	4.59 ± 0.75
Words Transcribed	47.57 ± 6.81	45.73 ± 8.39
Redundancy	1.82 ± 2.48	3.59 ± 3.24
Participant Utility	14.45 ± 7.67	10.61 ± 8.10
Agent Utility	17.38 ± 5.87	16.40 ± 8.63
Social Utility	31.83 ± 11.84	27.01 ± 13.67

respectively. Additionally, the standard deviation of effort levels varies between 0.13 and 0.17. This means that negative past experience resulted in more variability in effort levels of the participants compared to positive past experience.

Quartile distribution of effort levels indicates that more participants were inclined to complete more than half of the team task with negative past experience condition whereas more participants delivered half of the task with positive past experience condition. These differences demonstrate that the participants having negative past experience could not rely on the agent teammate as much as participants with positive past experience did for achieving team goals. It is likely that their negative past experience in the first game makes them play the game with increased caution.

5.3 Performance Analysis

The average cumulative outcomes of the second game for the two conditions are summarized in Table 1. The number of goals achieved and words transcribed in the positive past experience condition were marginally greater than those in the negative past experience condition ($F(1, 198) = 3.29,\ p < 0.1$; $F(1, 198) = 3.22,\ p < 0.1$). This is because the number of teams which failed to achieve their goal was higher for negative past experience compared to positive past experience over the course of the game. The sequence of the number of teams failed in five interactions are $(10, 6, 7, 4, 5)$ for positive past experience and $(21, 30, 29, 44, 32)$ for negative past experience.

Between Condition: Figure 3 presents the utility distribution in each interaction for positive and negative past experiences. Figure 3(a) depicts that the difference in social utility between the two conditions is negligible in the first two interactions. However, positive past experience ($M_3 = 6.97,\ SD_3 = 1.78$; $M_4 = 7.87,\ SD_4 = 3.87$; $M_5 = 8.74,\ SD_5 = 3.87$) generated significantly higher social utility than negative past experience ($M_3 = 5.79,\ SD_3 = 4.03$; $M_4 = 6.59,\ SD_4 = 5.08$; $M_5 = 7.07,\ SD_5 = 7.14$) in the third ($F(1, 198) = 7.15,\ p < 0.01$), fourth ($F(1, 198) = 3.99,\ p < 0.05$), and fifth ($F(1, 198) = 3.42,\ p < 0.1$) interactions.

Figure 3(b) shows participant utilities for the two conditions. Positive past experience led to higher participant utility throughout the game. Additionally, positive past experience led to significantly higher participant utility ($M_3 = 3.18,\ SD_3 = 1.35$; $M_4 = 3.58,\ SD_4 = 2.36$; $M_5 = 4.06,\ SD_5 = 2.85$) compared to negative past experience ($M_3 = 2.34,\ SD_3 = 2.17$; $M_4 = 2.59,\ SD_4 = 2.58$; $M_5 = 2.66,\ SD_5 = 3.65$) in the third ($F(1, 198) = 10.84,\ p < 0.01$), fourth ($F(1, 198) = 7.99,\ p < 0.01$), and fifth ($F(1, 198) = 9.03,\ p < 0.01$) interactions. The difference in agent utility between the two conditions is not significant over the course of the game as shown in Figure 3(c). This implies that the differences in social utility between the two conditions arise from participant utilities rather than agent utilities.

6 DISCUSSION

Previous research has emphasized the importance of past experience on people's future decision making and trust behavior [6, 10, 15, 28]. This study investigates how past experience affects human trust and effort levels in subsequent interactions with other agent teammates. Our findings suggest that past experience with an agent significantly affects the initial trust in another agent teammate in future interactions. However, the influence of past experience on trust behavior is likely to diminish over time. Furthermore, effort levels and utility distribution significantly differed between positive and negative past experiences. Negative past experience increased the participants' tendency to complete greater portions of the team task whereas positive past experience led the participants to expend significantly less effort, which reduced redundancy in team efforts. Hence, the latter engendered significantly higher participant and social utility.

Past experiences may affect various antecedents of trust including emotions, expertise, and attitudes [23, 25]. Among these, the variability of expertise between groups was avoided by comparing the games of the same order, i.e., both groups had the required expertise to complete assigned tasks. We believe that positive and negative past experiences altered the participants' emotions and attitudes towards agent teammates. The reason for the latter is that before the second game (in the GoT framework), the first game is the only similar past experience the participants ever had. Therefore, it is likely that their attitudes towards agent teammates were affected by their experience in the first game.

Trust: Trust in agent teammates is built on the expectations of the participants. Violation of trust nearly always causes the experience of negative feelings, such as distress, anger, and disappointment. On the contrary, fulfillment of trust leads to positive feelings, such as happiness, enthusiasm, and alertness. The results demonstrate that participants having positive past experience led to more initial trust than those having negative past experience (see Figure 1). **Hypotheses 1** and **2** are thereby supported. Our findings are in accordance with those of previous studies [14, 24, 28]. It could be argued that there are other factors that differ between the two conditions and can also affect the outcomes. Since the purpose of this particular study is to understand the relation between past experience (positive or negative) and trust in future interactions, the effects on other antecedents of trust can be investigated in future research.

Our results show that the difference in trust between positive and negative past experiences is initially significant and then gradually declines over the course of the game. Trust is a dynamic behavioral feature that is adjusted in every encounter of two or more entities based on their expectations and goals. Naturally, the weight of initial learned trust from past experience on the overall trust declines with successive interactions with the present teammate [23]. This finding is consistent with the findings of Stokes et al. [38] who demonstrated that the mood of an individual has a significant impact on initial trust formation and this impact diminishes as time and experience with the present automation increase.

Effort Level: Another key finding is that past experience affected human behavior in future interactions with another agent teammate. The results show that past experience significantly influenced the participants' effort levels in the second, fourth, and fifth interactions (see Figure 2(a)). One possible explanation for the difference is that past experiences of untrustworthiness may lead people to be more cautious. Therefore, the participants having negative past experience could not easily rely on their agent teammate, despite agent's contributions to teamwork. As a result, they continued to deliver significantly more work to minimize the risk of failing to complete the team goal. These findings support **Hypothesis 3**. What is surprising is that contrary to initial trust, initial effort levels by the participants did not significantly differ between positive and negative experiences. We expected the initial effort level to be influenced by past experience more than subsequent effort levels because the participants were completely uncertain about the trustworthiness of the agent teammate at the time of deciding initial contribution in the first interaction. This is an important issue and a further study is therefore warranted with more focus on the details of how the process influences trust and effort levels.

Team Performance: One of the objectives of this study is to examine whether trust in agent teammate and team performance are correlated. The results (see Table 1) show that the games in which the agent teammate was trusted more ended with a marginally greater number of achieved goals and words translated, significantly less redundancy, and, hence, significantly higher participant and social utility.

7 CONCLUSION

This study is an empirical investigation of the growth of human trust in human-agent teamwork in virtual environments without explicit communication. The novel aspect of this study that distinguishes it from previous work is that human and agent teammates have the same level of autonomy in a team. Key challenges arise from the uncertain and diverse nature of partner trustworthiness and the dynamic environment where a static allocation of tasks to team members or prior coordination is not possible due to the immediacy of team tasks, the impracticality of prior planning or limited communication.

We introduced a formal team game, the Game of Trust, and argue for its usefulness for studying human trust development for agent teammates over repeated interactions. We examined how past experience with agents affects human trust attitudes towards future agent teammates. Empirical findings show that positive past experience, e.g., prior interactions with a trustworthy agent teammate, led to significantly higher initial trust compared to negative past experience, e.g., prior interactions with an untrustworthy agent teammate. However, the impact of past experience diminished with the increase in experience with the present teammate. The effects of past experience are not limited to trust attitude. Positive (negative) past experience fosters (hinders) participants' reliance on agent teammates towards achieving team goals. As a result, positive past experience engenders an increase in reliance on agent teammates and concomitantly a reduction in redundant work, which improves team efficiency. These findings point out how influential past experiences can be on human-agent teamwork and provide lessons for agents to consider past experiences of potential human teammates while making interaction decisions.

Our future research priority is to study human-agent teamwork with complex tasks in ad-hoc scenarios. Such complex tasks comprise of subtasks that require different abilities as is experienced in many real-life teamwork scenarios. Furthermore, some of the subtasks may be dependent on others. Such ad-hoc scenarios are particularly challenging and interesting because humans and agents neither know each other's abilities regarding different task types nor the alignment of their own and teammate's abilities.

ACKNOWLEDGMENTS

This work is supported by the University of Tulsa, Bellwether (doctoral) Fellowship and Graduate Student Research Program.

REFERENCES

[1] Dimitrios Antos, Celso de Melo, Jonathan Gratch, and Barbara Grosz. 2011. The influence of emotion expression on perceptions of trustworthiness in negotiation. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (AAAIŠ11)*. San Francisco, CA, 772–778.

[2] Benoit A. Aubert and Barbara L. Kelsey. 2003. Further Understanding of Trust and Performance in Virtual Teams. *Small Group Research* 34, 5 (October 2003), 575–618.
[3] Robert Axelrod. 1984. *The Evolution of Cooperation.* Basic Books.
[4] Gary Bente, Odile Baptist, and Haug Leuschner. 2012. To buy or not to buy: Influence of seller photos and reputation on buyer trust and purchase behavior. *International Journal of Human-Computer Studies* 70, 1 (2012), 1–13.
[5] Moshe Bitan, Ya'akov Gal, Sarit Kraus, Elad Dokow, and Amos Azaria. 2013. Social Rankings in Human-Computer Committees. In *AAAI Conference on Artificial Intelligence.* 116–122.
[6] Grant Blank and William H. Dutton. 2012. Age and Trust in the Internet: The Centrality of Experience and Attitudes Toward Technology in Britain. *Social Science Computer Review* 30, 2 (2012), 135–151.
[7] Cody Buntain and Jennifer Golbeck. 2014. Trust Transfer Between Contexts.
[8] Christiano Castelfranchi and Rino Falcone. 2010. *Trust Theory: A Socio-Cognitive and Computational Model.* John Wiley & Sons.
[9] E. Chan and P. Vorderer. 2006. Massively multiplayer online games. In *Playing computer games: Motives, responses, and consequences*, P. Vorderer and J. Bryant (Eds.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 77–Ŭ88.
[10] Ying-Hueih Chen, Shu-Hua Chien, Jyh-Jeng Wu, and Pei-Yin Tsai. 2010. Impact of Signals and Experience on Trust and Trusting Behavior. *Cyberpsychology, Behavior, and Social Networking* 13, 5 (October 2010), 539–546.
[11] Celso de Melo, Peter Carnevale, and Jonathan Gratch. 2014. Social Categorization and Cooperation between Humans and Computers. In *The Annual Meeting of The Cognitive Science Society (CogSci'14).* 2109–2114.
[12] Ewart J. de Visser, Frank Krueger, Patrick McKnight, Steven Scheid, Melissa Smith, Stephanie Chalk, and Raja Parasuraman. 2012. The World is not Enough: Trust in Cognitive Agents. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56, 1 (September 2012), 263–267.
[13] Hongying Du and M.N. Huhns. 2013. Determining the Effect of Personality Types on Human-Agent Interactions. In *Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on.* 239–244.
[14] Jennifer R. Dunn and Maurice E Schweitzer. 2005. Feeling and Believing: The Influence of Emotion on Trust. *Journal of Personality and Social Psychology* 88, 5 (May 2005), 736–748.
[15] William H. Dutton and Adrian Shepherd. 2006. Trust in the Internet as an experience technology. *Information, Communication & Society* 9, 4 (November 2006), 433–451.
[16] Rino Falcone and Cristiano Castelfranchi. 2004. Trust Dynamics: How Trust is Influenced By Direct Experiences and By Trust Itself. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems.* IEEE Computer Society, Washington, DC, USA, 740–747.
[17] Xiaocong Fan, Sooyoung Oh, Michael McNeese, John Yen, Haydee Cuevas, Laura Strater, and Mica R. Endsley. 2008. The Influence of Agent Reliability on Trust in Human-agent Collaboration. In *Proceedings of the 15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction.* ACM, New York, NY, USA, 1–8.
[18] Ernst Fehr and Klaus M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* 114, 3 (August 1999), 817–868.
[19] Mark A. Fuller, Mark A. Serva, and John "Skip" Benamati. 2007. Seeing Is Believing: The Transitory Influence of Reputation Information on E-Commerce Trust and Decision Making. *Decision Sciences* 38, 4 (2007), 675–699.
[20] Alyssa Glass, Deborah L. McGuinness, and Michael Wolverton. 2008. Toward Establishing Trust in Adaptive Agents. In *Proceedings of the 13th International Conference on Intelligent User Interfaces.* ACM, New York, NY, USA, 227–236.
[21] Galit Haim, Yaakov Kobi Gal, Sarit Kraus, and Michele Gelfand. 2012. A Cultural Sensitive Agent for Human-Computer Negotiation. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (Eds.). International Foundation for Autonomous Agents and Multiagent Systems, Valencia, Spain, 451–458.
[22] Nader Hanna and Deborah Richards. 2014. "Building a Bridge": Communication, Trust and Commitment in Human-Intelligent Virtual Agent Teams. In *HAIDM Workshop at the Autonomous Agents and Multiagent Systems (AAMAS'14).*
[23] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors* 57, 3 (May 2015), 407–434.
[24] M. Hoogendoorn, S.W. Jaffry, P.P. van Maanen, and J. Treur. 2013. Modelling Biased Human Trust Dynamics. *Web Intelligence and Agent Systems Journal* 11, 1 (April 2013), 21–40.
[25] Gareth R. Jones and Jennifer M. George. 1998. The Experience and Evolution of Trust: Implications for Cooperation and Teamwork. *The Academy of Management Review* 23, 3 (July 1998).
[26] Aniket Kittur, Jeffrey V Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. 2013. The future of crowd work. In *Proceedings of the 2013 conference on Computer supported cooperative work.* ACM, 1301–1318.
[27] Sherrie Y. X. Komiak and Izak Benbasat. 2006. The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents. *MIS Quarterly* 30, 4 (December 2006), 941–960.
[28] Dietrich Manzey, Juliane Reichenbach, and Linda Onnasch. 2012. Human Performance Consequences of Automated Decision Aids: The Impact of Degree of Automation and System Experience. *Journal of Cognitive Engineering and Decision Making* 6 (January 2012), 57–87.
[29] Roger C. Mayer, James H. Davis, and F. David Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20, 3 (July 1995), 709–734.
[30] Daniel J. McAllister. 1995. Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal* 38, 1 (February 1995), 24–Ŭ59.
[31] Stephanie M. Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 2013. I Trust It, but I DonŠt Know Why Effects of Implicit Attitudes Toward Automation on Trust in an Automated System. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 55, 3 (June 2013), 520–534.
[32] Tim Robert Merritt, Kian Boon Tan, Christopher Ong, Aswin Thomas Abraham, Teong Leong Chuah, and Kevin McGee. 2011. Are artificial team-mates scapegoats in computer games?. In *Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work, CSCW.* ACM, 685–688.
[33] Florian Nothdurft, Tobias Heinroth, and Wolfgang Minker. 2013. The Impact of Explanation Dialogues on Human-Computer Trust. In *Human-Computer Interaction. Users and Contexts of Use*, Masaaki Kurosu (Ed.). Springer Berlin Heidelberg, 59–67.
[34] Christopher Ong, Kevin McGee, and Teong Leong Chuah. 2012. Closing the human-AI Team-mate Gap: How Changes to Displayed Information Impact Player Behavior Towards Computer Teammates. In *Proceedings of the 24th Australian Computer-Human Interaction Conference.* ACM, New York, NY, USA, 433–439.
[35] R. Pak, N. Fink, M. Price, B. Bass, and L. Sturre. 2012. Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics* 55, 9 (July 2012), 1059–1072.
[36] Paul Robinette, Alan R. Wagner, and Ayanna M. Howard. 2013. Building and Maintaining Trust Between Humans and Guidance Robots in an Emergency. In *AAAI Spring Symposium: Trust and Autonomous Systems.* Stanford, CA, 78–83.
[37] Sandip Sen. 2013. A Comprehensive Approach to Trust Management. In *Proceedings of the Twelfth International Conference on Autonomous Agents and Multiagent Systems.* Saint Paul, Minnesota, USA, 797–800.
[38] C.K. Stokes, J.B. Lyons, K. Littlejohn, J. Natarian, E. Case, and N. Speranza. 2010. Accounting for the human in cyberspace: Effects of mood on trust in automation. In *Collaborative Technologies and Systems (CTS), 2010 International Symposium on.* 180–187.
[39] Piotr Sztompka. 1999. *Trust: A Sociological Theory.* Cambridge University Press, Cambridge, UK.
[40] A. van Wissen, Y. Gal, B.A. Kamphorst, and M. V. Dignum. 2012. Human-agent teamwork in dynamic environments. *Computers in Human Behavior* 28, 1 (2012), 23–33.
[41] Arlette van Wissen, Jurriaan van Diggelen, and Virginia Dignum. 2009. The Effects of Cooperative Agent Behavior on Human Cooperativeness. In *Proceedings of eighth International Confefence on Autonomous Agents and Multiagent Systems (AAMAS 2009).* 1179–1180.
[42] Frank Verberne, Jaap Ham, and Cees J. H. Midden. 2012. Trust in Smart Systems: Sharing Driving Goals and Giving Information to Increase Trustworthiness and Acceptability of Smart Systems in Cars. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54, 5 (May 2012), 799–810.
[43] Frank M. F. Verberne, Jaap Ham, and Cees J. H. Midden. 2014. Familiar faces: Trust in a facially similar agent. In *HAIDM Workshop at the Autonomous Agents and Multiagent Systems (AAMAS'14).*