Understanding The Influences of Past Experience on Trust in Human-Agent Teamwork

FEYZA MERVE HAFIZOĞLU, İstanbul Şehir University, Turkey
SANDIP SEN, The University of Tulsa, USA

People use the knowledge acquired from past experiences in assessing the trustworthiness of a trustee. In a time where the agents are being increasingly accepted as *partners* in collaborative efforts and activities, it is critical to understand all aspects of human trust development in agent partners. 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 reliance on agent teammates. To do so, we developed a 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; lack of past experience leads to higher trust levels compared to positive past experience; positive (negative) past experience facilitates (hinders) reliance on agent teammate; the relationship between trust in and reliance on agent teammates is not always correlated. 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**.

Additional Key Words and Phrases: Human-agent teamwork, trust, reliance, past experience

ACM Reference Format:
Feyza Merve Hafızoğlu and Sandip Sen. 2019. Understanding The Influences of Past Experience on Trust in Human-Agent Teamwork. *ACM Trans. Internet Technol.* 1, 1, Article 1 (January 2019), 22 pages. https://doi.org/0000001.0000001

1 INTRODUCTION

Experience is one of the fundamental influences that help shape our behavior. For example, we use the knowledge acquired from past experiences in assessing the trustworthiness of a trustee, such as a person, an organization, or a computer agent [6, 10, 32]. Past experiences have an impact on cognitive-based trust as well as affective-based trust [34]. In particular, past experience can alter a variety of aspects of our lives, including our expertise, attitudes, expectations, feelings, and so on. Thus, past experience has a significant influence on people's trust relationships.

Authors' addresses: Feyza Merve Hafızoğlu, İstanbul Şehir University, Turgut Özal Blv, No: 21, Kartal, İstanbul, Turkey, feyzahafizoglu@sehir.edu.tr; Sandip Sen, The University of Tulsa, 800 S. Tucker Drive, Tulsa, Oklahoma, 74104, USA, sandip@utulsa.edu.

1533-5399/2019/1-ART1 $15.00
https://doi.org/0000001.0000001

Trust as social glue plays a critical role in inducing cooperative behavior among individuals and within groups [3, 27, 33, 34]. The role of trust is not limited to human-human interactions: trust also shapes the way people engage with technology [8, 19, 27, 45]. As autonomous agents are being increasingly accepted as human *partners* in collaborative efforts and activities, establishing people's trust in agent teammates has been recognized as an important objective in the multi-agent research community. In order to achieve fluid interactions between humans and agents, we must have a clear understanding of the formative process of human trust in agent partners. Clarifying and understanding this process can help us to develop more effective and useful agent applications in domains where humans recognize agents as *autonomous* and *effective* partners and have to rely on agents as "peer" level team members. Research on human trust in agent teammates is, despite the increasing body of work recently, still limited and we need more evidence from diverse scenarios to better understand the underlying factors of human trust in agent teammates.

The primary objective of this study is 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.

The specific objective of this study is to investigate the effects of *past experience* on human trust in and reliance on agent teammates. *Past experience* in this context refers to prior interactions with other agents in the same or similar situations. People's past experience can affect the perceived trustworthiness of agent teammates and the outcome of teamwork [17].

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 in and rely on an agent teammate? What are the effects of experience with an untrustworthy agent on the subsequent interactions with other agent teammates? What are the differences between presence and lack of experience with respect to trust and reliance?

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 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 an extensive set of 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 gives a brief overview of the recent history of human trust in automation and agents. 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 the experiments and discuss the empirical findings in Section 6. Finally, Section 7 provides a summary of the study and the directions for future research.

2 RELATED WORK

The significance of trust in human-agent interactions has been well-acknowledged in literature [12, 18, 21, 26, 46]. 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, 40], personality [13], culture [23], mood [31, 44], attitude [35], and past experience [6, 10, 32].

Previous research demonstrates that the effect of past experience on human trust behavior towards technology differs between context. Manzey et al. [32] discover that negative past experience causes operators to have reduced trust in automation, while Sauer et al. [42] found that over-reliance increases when the participants have training with an unreliable automation compared to a fully reliable automation. Chen et al. [10] found that past experience with websites enhanced the perceived ability, integrity, benevolence, and predictability of e-vendors. On the contrary, Hergeth et al. [25] demonstrated that prior familiarization with automation would decrease initial trust compared with no prior familiarization in automated driving domain. Dutton et al. [15] suggest that positive (negative) past experiences led to trust more (less), however, this influence diminishes over time. Relevant to positive/negative past experience, it has been demonstrated that positive (negative) emotions [14] and mood [31] have positive (negative) influence on trust based on certain situational cues. 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 [47] and reliability [18], improves trust and facilitates the collaboration between parties. In contrast, negative behavior, such as defection [46] and deception [47], 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 [21, 38, 48]. Furthermore, familiarity and personalization of agents have been shown to positively influence human trust [29, 49].

In addition to agent behavior, researchers have investigated the effects of different agent representations, such as avatars and robots [1, 11, 41, 49], and the effects of external factors, such as information representation [4] and reputation [20].

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 [24], negotiation [46], giving advice [12, 44], providing recommendations [29], and physical interaction [41].

Recently, new environments, that enable group activities or collaboration between humans and agents, have been emerging, such as crowd-work with complex tasks [28] 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 aspect of human trust. In a study on human-agent teamwork without explicit coordination, Merritt et al. [36] examined the blame behavior for team failures. In another study, Ong et al. [39] 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, 47, 49]; focusing on teamwork environments in which there is neither explicit communication between human and agent (as in [24, 46]) nor agents embodied in physical forms, such as robots (as in [1, 11, 24, 49]); exploring repeated, in contrast to one-shot [48], interactions in fixed rather than dynamic teams [46]; providing real team tasks for evaluating human-agent teamwork rather than the standard artificial environments [1, 11, 24, 46–48]. 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 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* [43]. 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. The 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 size, 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 the two players 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. Since these tasks are just "decoys" that we use to evaluate growth of human trust from repeated interactions, their completion is of no intrinsic value to us, and we are not studying optimal task allocation. We simply count the number of words accurately transcribed and give credit even when the team members transcribe overlapping word sets. In this domain, the term *task size* refers to the number of words to transcribe, i.e., the number of subtasks, in an interaction.

The purpose of the transcription task is to mimic a real teamwork environment where the human players 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 human players 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 [50] over the length of the transcribed word. This is done to ensure a minimum quality of human player 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

[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.

teamwork instances. We expect that interacting with a (an) trustworthy (untrustworthy) agent teammate is most likely to inspire positive (negative) feelings. Specifically, it is demonstrated that happiness significantly improves trust whereas anger significantly reduces it [14]. This finding is further supported by the previous studies on the relation between the essence of the experience, and resulting trust behavior [10, 32]. Therefore, we posit that positive (negative) past experience leads participants to perceive their agent teammate more (less) trustworthy:

HYPOTHESIS 1. Positive past experience leads to higher initial trust in future agent teammates compared to negative past experience.

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

HYPOTHESIS 2. Past experience with an agent teammate affects effort levels by the participants in subsequent interactions with other agent teammates.

People tend to believe that computers perform tasks flawlessly and trust them not to make mistakes. In other words, they ignore the possibility of less than optimal software performance caused by incomplete or inaccurate information, unpredictable domain events, or even software errors. Accordingly, previous studies provide evidence that people are biased to a priori perceive automation to be trustworthy [16, 26]. However, experiencing sub-par system performance or failures can immediately undermine this trust bias and users will also subsequently have lower trust levels or expectations. It can even happen that naive, inflated expectations are not met by generally positive, yet not outstanding interaction outcomes with partners. We therefore expect that participants without past experience would put higher initial trust in agent teammates.

HYPOTHESIS 3. The participants without past experience are likely to have higher initial trust in agent teammates compared to the participants with past experience.

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 agent 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 subtasks completed by the Trustworthy agent in i^{th} interaction is

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

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

where t_i is the team task size of i^{th} interaction, w_h^{i-1} is the number of subtasks completed by the human player 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 agent player* that is neither a dummy player, 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 the participants believe their teammate is inclined to exploit them whenever there is a chance, e.g., reducing its efforts when a participant consistently delivers more than a fair share. The untrustworthy player makes at least one unfair choice in a game. The number of subtasks 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 i^{th} interaction, Δ^i, is stochastically incremented. Therefore, its effort is monotonically non-increasing and decreases occasionally. 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 describes the task size choice function of the Untrustworthy agent player. 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 would be to increase its effort. Accordingly, the Untrustworthy agent completes half of the team task, i.e., $t_i/2$ subtasks, if the recent three interactions were failures (lines 1-2). Likewise, it completes half of the team task, i.e., $t_i/2$ subtasks, or half of the team task minus one, i.e., $t_i/2 - 1$ subtasks, if the last two interactions were failures by setting a random value (either zero or one) to Δ (lines 3-5).

The third condition (lines 6-8) ensures that the Untrustworthy agent exhibits untrustworthy behavior at least once in a game. If the value of Δ has not been incremented after three interactions, that means the Untrustworthy agent has delivered half of the team task so far. In that case, the agent is forced to deliver less than the fair share by incrementing the value of Δ.

In the else condition (line 9), when the first three conditions do not hold, *Bernoulli distribution* is used to determine whether the value of Δ will be incremented (lines 18-21). In an interaction, the base value of the *Bernoulli distribution* parameter, p, is initialized with p_{min}, a global variable that is meant to be the maximum base value in a game (line 10). 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 (lines 15-17). 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, i.e., extremely lower values of task size choice, the value of minimum probability to increment Δ, p_{min}, is subsequently reduced by 0.05 (line 20). Finally, the individual task size is computed as half of the team task minus Δ, i.e., $t_i/2 - \Delta$ (line 23).

4.1.3 Learner Agent. *Learner agent* is trained offline to predict human player's task choices by utilizing a linear regression model that is trained with the data collected from previous experimentation [22]. It delivers half of the team task 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.

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

```
if nfailures ≥ 3 then
    Δ ← 0
else if nfailures ≥ 2 then
    Δ ← x // random number x ∈ [0,1]
end
else if i > 3 and Δ = 0 then
    Δ ← 1
end
else
    p ← p_min
    ϵ ← 0
    if i > 1 then
        ϵ ← w_h^{i-1} / t_{i-1} − 0.5
    end
    if ϵ > 0 then
        p ← p + ϵ /* Increase the probability to increment Δ */
    end
    if rand(0, 1) < p then
        Δ ← Δ + 1
        p_min ← p_min − 0.05 /* Higher the value of Δ, lower the probability to increment Δ */
    end
end
w^i_Untrustworthy ← t_i / 2 − Δ
return w^i_Untrustworthy
```

However, given the irrational, unpredictable, and "noisy" behavior of human players, it is a challenge to develop a learning agent player that can produce optimal social utility over repeated interactions. This is particularly true given that any adaptation by agents can elicit responsive adaptation by the human, which significantly complicates the task of the Learner agent.

4.2 Experimental Setup

4.2.1 Game Configuration. The number of interactions in a game is five (as in [5, 7, 39]), which is short enough to avoid the participants becoming bored while providing some experience that allows the team members to adapt to teammates with predictable behavior. The team is assigned a team task, consisting of several subtasks, in each interaction. The size of the team task, i.e., the number of subtasks (each subtask in our domain involves the transcription of a word), 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 subtask are set to 1 and 1.75, respectively. The players are allowed to choose a task size, i.e., the number of subtasks, between one and the size of team task minus one.

4.2.2 Experimentation. In order to evaluate the effect of past experience, we performed two experiments: one for comparing the consequences of positive and negative past experience and one for comparing the effects of the presence of past experience (positive or negative) with no past experience. Each experiment consists of two games. The participants were instructed that they will play with a new computer player at the beginning of each game.

Experiment 1. Positive vs. Negative Past Experience: The first of the two games is for providing the participants 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 the type of 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 the participants having positive (negative) past experience, respectively. 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[2] was their past experience; the order of the game and the agent player were the same for both groups.

Experiment 2. No vs. Positive/Negative Past Experience: Similar to Experiment 1, the first and second games in this experiment resemble the past experience with an agent player and repeated interactions with a new agent player, respectively. Additionally, the first game is used to investigate the no past experience condition. We experimented with two groups of participants based on the type of experience, *G3* and *G4*, with the following associated teammate orderings:

G3: Trustworthy Agent, Untrustworthy Agent;
G4: Untrustworthy Agent, Trustworthy Agent.

We investigate the consequences of positive past experience by contrasting with having no past experience. To do so, the first game played with the Untrustworthy agent by group G4 (no past experience) and the second game played with the same agent by group G3 (positive past experience) are compared. Likewise, to evaluate the effect of negative past experience, the first game played with the Trustworthy agent by group G3 (no past experience) and the second game played with the same agent by group G4 (negative past experience) are compared. Within each comparison (no vs. positive and no vs. negative), the participants played with the same agent player and hence the only relevant difference between the two groups is the presence of past experience.

4.2.3 Trust Survey. The game includes a short survey on trust to assess the participants' perceived trustworthiness and fairness of their agent teammates. The participants completed this survey after the first, third, and fifth interactions of a game (similar to [44]) after they were shown the outcome

[2] Two different sets of participants were sampled from the same population to play in the two groups, as it was not feasible to utilize the same set of participants to test both positive and negative experiences.

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,
(3) My teammate works responsibly for accomplishing the team task.

In an interaction, a participant's trust in her agent teammate is computed as the average of the responses to these three items.

4.2.4 Metrics. Three essential metrics are adopted to assess the influences of past experience in our analysis.

(1) *Trust level* in an agent teammate is the average of the participants' average responses to the three items in the trust survey and has a value in the range $[1, 5]$.
(2) *Effort level* is the portion of the total subtasks completed by the player, i.e., the fraction of individual task size over the team task size, and has a value in the range $[0, 1)$. This metric reveals how participants' reliance is affected: higher the effort level by the participant, lower the participant's reliance on agent player is, and vice versa.
(3) *Cumulative game results* is a collection of the outcomes from all five interactions in a game: (1) the total number of team goals achieved (successful interaction) that has a value in the range $[1, 5]$, (2) the total number of subtasks, i.e., words transcribed, by the team that has a value in the range $[5, 90]$, (3) the average number of words transcribed redundantly (the number of excess subtasks performed) that has a value in the range $[0, 5]$, and (4) the total participant/agent/social utilities. These metrics are aimed to analyze the relationship between past experience and team performance.

4.2.5 Participants. We recruited 339 participants through Amazon Mechanical Turk[3]. Data of 24 participants was eliminated due to insufficient attention[4]. There were 98, 102, 55, and 60 participants in groups G1, G2, G3, and G4, respectively. Approximately 50% of the participants were female. Age distribution was as follows: 18-24 years, 18%; 25-34 years, 47%; 35-44 years, 20%; 45-54 years, 10%; 55-64 years, 4%; and 65 years or older, 1%. The distribution of education levels was as follows: high school degree, 13%; some college experience, 29%; associate's degree, 9%; bachelor's degree, 34%; and graduate degree, 14%; and PhD, 1%. The ethnicity distribution was as follows: White, 79%; Hispanic-Latino, 4%; African-American, 7%; Asian, 7%; and other ethnicity, 3%.

5 RESULTS

This section presents *trust level*, *effort level*, and *team performance* analysis. The average of these metrics among the participants is reported. The mean (M) and standard deviation (SD) values are provided and a subscript is used (M_i, SD_i) to indicate the interaction number if necessary. One-way ANOVA analysis is used to test the statistical significance.

5.1 Trust Analysis

Positive vs. Negative Past Experience: In order to compare the effects of negative and positive past experience on trust in future agent teammates, trust levels are computed from the experimental

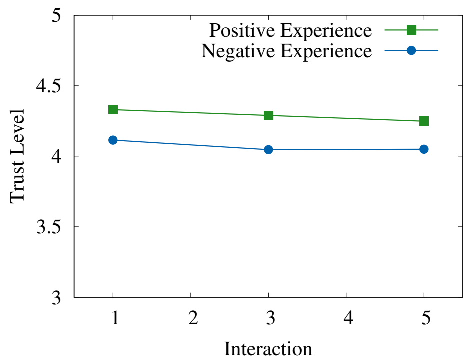
[3]http://www.mturk.com/

[4]For monitoring the participants' attention, trust survey has one bogus and one consistency item, which have a similar or opposite meaning with one of the trust items listed above. If a participant provides an invalid response to a bogus item, their study is terminated and they cannot participate again. Consistency items were used to determine the level of attention and the data of participants who did not pay enough attention was filtered.

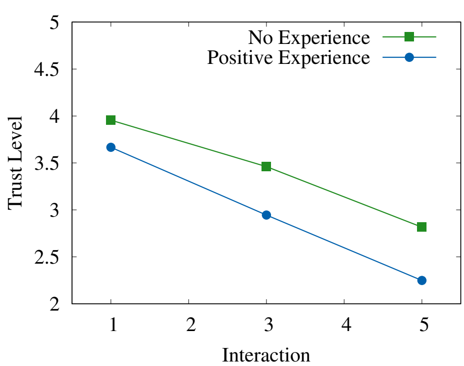
data that is collected from the second game played with the Learner agent by G1 and G2. Figure 1(a) depicts the participants' trust in the Learner agent after the first, third, and fifth interactions for the two conditions.

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

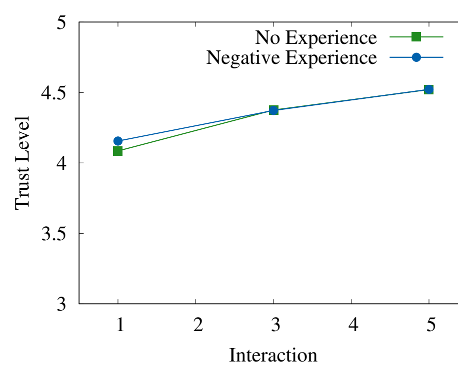
Between Condition: The participants differentiated between the conditions with respect to their trust in the agent teammate. Higher levels of trust for positive past experience condition demonstrate that positive experience facilitated the trust formation process. In particular, initial trust is significantly 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 the present experience.



(a) Trust in the Learner agent for positive and negative past experiences



(b) Trust in the Untrustworthy agent for no and positive past experiences



(c) Trust in the Trustworthy agent for no and negative past experiences

Fig. 1. Trust levels for no past experience compared to positive/negative past experiences

Figure 1(b) and 1(c) depict the participants' trust in their agent teammates for no, positive, and negative past experiences in *Experiment 2*.

No vs. Positive Past Experience: Figure 1(b) shows how trust in the Untrustworthy agent varies throughout the game for no and positive past experience.

Within Condition: As shown in Figure 1(b), trust in the Untrustworthy agent significantly worsens over the interactions in both no past experience ($F(2, 162) = 14.96,\ p < 0.001$) and positive past experience ($F(2, 162) = 19.19,\ p < 0.001$) conditions. This decline is expected due to the unreliable nature of the Untrustworthy agent making unfair task choices.

Between Condition: Trust in the Untrustworthy agent is significantly higher in no past experience condition ($M_3 = 3.46,\ SD_3 = 1.28; M_5 = 2.82,\ SD_5 = 1.15$) in comparison with positive past experience condition ($M_3 = 2.95,\ SD_3 = 1.41; M_5 = 2.25,\ SD_5 = 1.03$) after the third ($F(1, 118) = 4.21,\ p < 0.05$) and fifth ($F(1, 118) = 6.90,\ p < 0.01$) interactions. Despite playing with the same agent teammate in both conditions, the continuing difference in resulting trust levels reveals that having or not having past experience has a lasting impact on the participants' trust in agent teammates. Specifically, no past experience lead to greater trust compared to having even positive experience.

No vs. Negative Past Experience: Figure 1(c) compares trust in the Trustworthy agent for no and negative past experience.

Within Condition: As shown in Figure 1(c), the participants increasingly trust in the Trustworthy agent in both no past experience ($F(2, 162) = 5.93,\ p < 0.01$) and negative past experience ($F(2, 162) = 5.36,\ p < 0.01$) conditions. This improvement is a reflection of the reliable behavior exhibited by the Trustworthy agent on the participants' trust.

Between Condition: Interestingly, no significant differences were found between the two conditions over interactions despite the slight difference in initial trust level. The participants with negative past experience greatly appreciated the performance of the Trustworthy agent. The resulting trust levels are equal to or slightly higher than those in no experience condition.

The above results suggest that current positive experience, e.g., collaborating with the Trustworthy agent, can compensate for the initial trust bias of users without prior experience. However, current negative experience will prolong the trust differential between participants with no experience and even positive past experience.

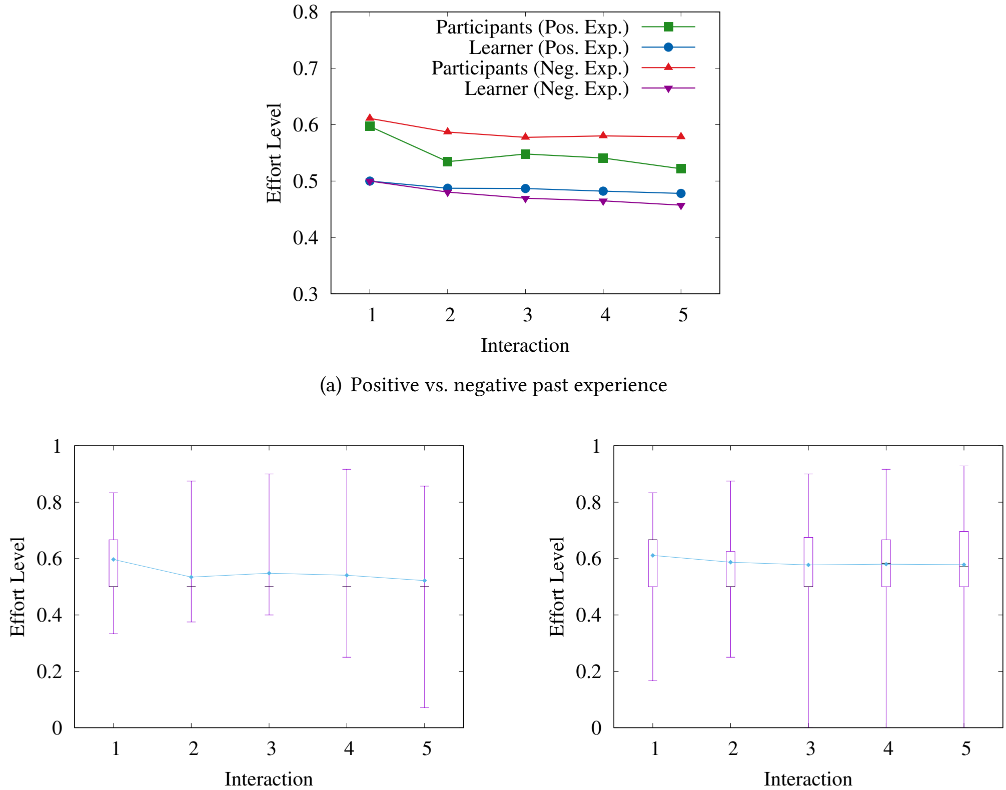
5.2 Effort Level Analysis

Positive vs. Negative Past Experience: Figure 2 presents the effort level distribution and the quartile distribution of effort levels for positive and negative past experience conditions in *Experiment 1*.

Figure 2(a) depicts the effort levels by the participants and the Learner agent for the two conditions.

Within Condition: 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, the 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 the 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, the initial effort levels by the participants were not significantly affected by past experience. In the second interaction, however, the 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



(a) Positive vs. negative past experience

(b) Positive past experience

(c) Negative past experience

Fig. 2. Effort levels by the participants and the Learner agent and quartile distribution of effort levels by participants for positive and negative past experiences

those for negative past experience ($M = 0.59,\ SD = 0.13$). Similarly, the effort levels for positive past experience ($M_4 = 0.54,\ SD_4 = 0.11; M_5 = 0.52,\ SD_5 = 0.12$)[5] 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 influence on the participants' effort levels as follows: negative past experience led the participants to make more cautious choices whereas positive past experience fostered more reliance on the agent teammate to complete the team task and, thereby, led to reduced effort levels.

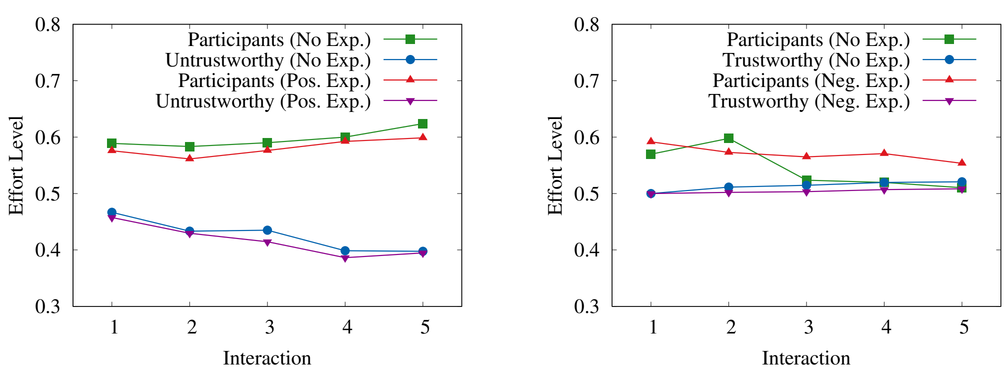
In Figures 2(b) and 2(c), we present the quartile distribution of effort levels by the participants for positive and negative past experiences. Figure 2(b) shows that positive past experience led the majority of the participants to deliver half of the work. In the first interaction, the upper quartile having a value of 0.67 illustrates that the 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

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

deviation of the effort levels varies between 0.8 and 0.12 in the five interactions, which indicates the homogeneity of the effort levels for positive past experience.

Figure 2(c) shows the quartile distribution of the effort levels for negative past experience. Initially, at least 50% of the participants put in more than 0.67 effort, i.e., 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 to 0.63 in the second interaction. In the subsequent interactions, their values increased to 0.57 and 0.70, respectively. Additionally, the standard deviation of the effort levels varies between 0.13 and 0.17. This demonstrates that negative past experience resulted in effort level variability compared to positive past experience.

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



(a) No vs. positive past experience

(b) No vs. negative past experience

Fig. 3. Effort levels by the participants and the agent teammates for no and positive/negative past experiences

Figure 3 depicts the effort level distribution[6] for no and positive/negative past experience conditions when playing with the Untrustworthy and Trustworthy agent in *Experiment 2*.

No vs. Positive Past Experience: Figure 3(a) shows the effort levels by the participants and the Untrustworthy agent for the two conditions.

Within condition: The participants without past experience slightly improved their effort level towards the end of the game. On the other hand, the participants having positive past experience initially reduced and then improved their effort levels ($F(4, 270) = 2.66,\ p < 0.05$). In both games, the participants began with an effort level close to 0.6, maintained it in the second interaction, and increased the effort levels in the rest of the game. For both conditions, the Untrustworthy agent occasionally reduced its effort level over interactions. There is a clear trend of increasing effort levels by the participants in response to decreasing effort levels by the Untrustworthy agent in

[6]The corresponding quartile distributions for no vs. positive past experience and no vs. negative are found to be very similar with each other. Therefore, they are not presented.

both games. In other words, the Untrustworthy agent adapts its performance based on the level of reliance by the participants: lower the reliance, lower the performance is.

Between condition: The effort levels by the participants having positive past experience are slightly lower than those in no past experience condition throughout the game. A very similar trend is observed in the effort levels by the Untrustworthy agent between the two conditions, i.e., slightly lower effort levels in positive past experience condition.

Positive past experience leads to slightly lower effort levels compared to no past experience over interactions. What is surprising and noteworthy is that the Untrustworthy agent is trusted significantly more by the participants without past experience as shown in Figure 1(b) and yet these participants still put slightly higher efforts compared to positive past experience condition. In other words, they can trust but cannot rely on the agent teammate to the degree to which they trust. This result reveals that there exist factors affecting reliance other than trust, i.e., being unfamiliar to a domain may cause people to act more cautiously.

No vs. Negative Past Experience: Figure 3(b) shows the effort level distribution for no and negative past experience when playing with the Trustworthy agent in *Experiment 2*.

Within condition: In no past experience condition, the participants' effort levels that are initially improved and then reduced, significantly vary ($F(4, 270) = 5.15,\ p < 0.001$), while the Trustworthy agent continuously increased its effort level over interactions ($F(4, 270) = 2.66,\ p < 0.05$). On the other hand, no significant variation is found in either the participants' or the Trustworthy agent's effort levels in the negative past experience condition.

Between condition: Initially, negative past experience causes slightly higher effort levels compared to no past experience condition. While there is a steady decline in effort levels of the participants having negative past experience, having no past experience lead to unstable rise and fall of effort levels in the first three interactions. The participants without past experience put significantly lower effort ($M_4 = 0.52,\ SD_4 = 0.09; M_5 = 0.51,\ SD_5 = 0.07$) than those in negative experience condition ($M_4 = 0.57,\ SD_4 = 0.14; M_5 = 0.55,\ SD_5 = 0.11$) after the fourth ($F(1, 118) = 5.33,\ p < 0.05$) and fifth ($F(1, 118) = 5.59,\ p < 0.05$) interactions. This result reveals that negative past experience did not allow the participants to rely on their agent teammate despite its trustworthy behavior. The effort levels by the Trustworthy agent are marginally ($p < 0.1$) lower in negative past experience condition compared to those in no past experience condition in the last four interactions. Though marginal, this difference is a result of the higher effort levels by the participants in negative past experience condition compared to no past experience condition, i.e., the Trustworthy agent adapting its performance based on the participants' reliance.

To summarize, no past experience leads to fluctuating initial behavior that is followed by reliance on trustworthy agent teammate, i.e., the participants were able to learn to trust in and rely on the Trustworthy agent. In contrast, negative past experience caused the participants to remain cautious throughout the game. They were not able to rely on their agent teammate that exhibits totally trustworthy behavior.

5.3 Performance Analysis

To analyse the relationship between past experience and team performance, the average cumulative outcomes from the games of our interest are presented below.

Table 1 provides a summary of the second games played with the Learner agent in *Experiment 1*. The number of goals achieved and words transcribed in positive past experience condition were marginally greater than those in 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

Table 1. Results of the games with the Learner agent 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

Table 2. Results of the games with the Untrustworthy agent for no and positive past experiences

	No Past Experience	Positive Past Experience
Goals Achieved	3.45 ± 1.05	3.27 ± 1.08
Words Transcribed	33.50 ± 10.90	31.24 ± 11.90
Redundancy	3.35 ± 3.41	2.07 ± 2.23
Participant Utility	−0.94 ± 7.55	−1.96 ± 7.70
Agent Utility	8.34 ± 10.60	6.84 ± 11.72
Social Utility	7.41 ± 16.92	4.88 ± 18.07

Table 3. Results of the games with the Trustworthy agent for no and negative past experiences

	No Past Experience	Negative Past Experience
Goals Achieved	4.67 ± 0.67	4.82 ± 0.50
Words Transcribed	46.95 ± 6.29	47.93 ± 5.93
Redundancy	3.11 ± 3.29	3.90 ± 5.07
Participant Utility	14.15 ± 4.66	13.41 ± 6.63
Agent Utility	15.31 ± 6.99	16.69 ± 5.73
Social Utility	29.46 ± 10.86	30.10 ± 11.20

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.

Table 2 presents the results of games played with the Untrustworthy agent for no past experience and positive past experience in *Experiment 2*. The number of goals achieved and words transcribed are slightly higher for no past experience. Interestingly, the redundancy for no past experience is significantly higher than that for positive past experience ($F(1, 118) = 5.19,\ p < 0.05$). As a result of significantly less redundant work, participant, agent, and social utilities are higher for no past experience.

Table 3 presents the cumulative game results played with the Trustworthy agent for no past experience and negative past experience in *Experiment 2*. No significant differences are found between the two conditions. The number of goals achieved and words transcribed are slightly higher for negative past experience. The redundancy is also slightly higher for negative past experience. This redundant work lead to lower participant utility for negative past experience.

6 DISCUSSION

Previous research has emphasized the importance of past experience on people's future decision making and trusting behavior [6, 10, 15, 32]. This study investigates how past experience affects human trust in and reliance on future agent teammates in subsequent interactions. Past experiences may affect various antecedents of trust including feelings/emotions, expertise, expectations, and attitudes [26, 27]. We consider three factors associated with past experience: feeling, expertise (experience), and expectations.

In the first experiment, the variability of expertise within GoT between the 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, expectations, and attitudes[7] towards agent teammates. Moreover, 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. The second experiment aims to assess the role of past experience by comparing lack of past experience with positive and negative past experience. Specifically, we compare the two games: one played without past experience and one played with positive/negative experience. Between the two groups, the former had no expertise in GoT, and hence no bias and expectations associated with GoT. On the contrary, the latter had expertise and expectations that were acquired through the previous game. Given that the games were played with the same agent player, the only difference is the presence/lack of past experience.

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. Interestingly, lack of past experience can result in greater trust compared to positive past experience. Furthermore, effort levels and utility distribution significantly differed between positive and negative past experiences. Negative past experience reduced reliance, i.e., increased the participants' tendency to complete greater portions of the team task, whereas positive past experience fostered reliance, i.e., led the participants to expend significantly less effort, which reduced redundancy in team efforts. Hence, the latter engendered significantly higher participant and social utility. Additionally, major differences are found in effort levels between lack of past experience and negative past experience.

Trust: Violation of trust nearly always causes the experience of negative feelings, such as distress, anger, and disappointment. On the contrary, the fulfillment of trust leads to positive feelings, such as happiness, enthusiasm, alertness, and so on. Accordingly, the consequences of positive (negative) past experience are two-fold: positive (negative) past experience leads to positive (negative) feelings and, besides, raise (drop) expectations. Trust in agent teammates is built on the expectations of the participants: greater expectations require greater performance by the trustee in order to trust and vice versa. The results show that positive past experience led to greater initial trust than negative past experience when playing with an adaptive agent player (see Figure 1(a)). It seems the effect of positive feelings outweighed the difficulty of fulfilling higher expectations. **Hypothesis 1** is thereby supported. Previous studies are consistent with with the present results and have demonstrated that positive (negative) experience [15, 32], emotions [14] and mood [31] have positive (negative) influence on trust based on situational cues. However, this outcome is contrary to that of Sauer et al. [42] who found that automation bias (over-reliance) [37] increases when the participants have training with an unreliable automation compared to a fully reliable automation. This inconsistency may be due to dissimilarity of domains and the level of autonomy (agent vs. automation). 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, we leave the investigation of the effects on other antecedents of trust for future research.

[7] Even though it is likely that their attitudes towards agent teammates were affected by their experience from the first game, we omitted the attitude factor in this study.

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 [26]. This finding is consistent with the previous studies [31, 44] that 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.

The second aim of this study was to investigate the effects of past experience with respect to lack of past experience. The results demonstrate that no past experience leads to significantly greater trust compared to positive past experience (see Figure 1(b)) when interacting with an untrustworthy agent player. Interestingly, having no past experience had a lasting impact on trust throughout the game. However, comparing no past experience with negative past experience, when interacting with a trustworthy agent player, made no significant difference (see Figure 1(c)). The combination of having positive past experience, i.e., positive feelings as well as high expectations, and collaboration with an untrustworthy agent teammate, i.e., poor performance, hinders trust formation. On the contrary, the combination of low expectations due to negative past experience and credible performance by a trustworthy agent teammate aids it. We posit that the former situation increased the difference in trust levels between no and positive past experience conditions, while the latter decreased the difference between no and negative past experiences. Taken together, these findings suggest that no past experience has a positive impact on trust formation. Equally important, the effect of past experience on expectations plays a significant role on trust. In particular, positive (negative) past experience raises (drops) expectations which be difficult (easy) to fulfill by future agent teammates. Hence our findings partially support **Hypothesis 3** and previous work [16, 26]. These results reflect those of Hergeth et al. [25] who report that prior familiarization with automation would decrease initial automation trust compared with no prior familiarization. Since this investigation is limited to agent teammates with certain behavior, the generalizability of these results is subject to certain limitations. Therefore, further experimentation with different agent teammate behavior should be performed to collect more evidence.

Reliance: Another key finding is that past experience affected participants' reliance strategy, i.e., effort level choices, in future interactions with another agent teammate. The results show that past experience significantly influenced the degree to which the agent teammates were relied upon in both experiments (see Figures 2(a) and 3(b)). 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 (see Figures 2(a) and 3(b)). As a result, they continued to deliver significantly more work to minimize the risk of failing to complete the team goal. On the other hand, positive past experience facilitates relying on agent teammates resulting with lower effort levels in both experiments. Comparing lack of past experience with positive and negative past experience, the results indicate that positive past experience facilitates reliance slightly better than no past experience does, while no past experience assures significantly better reliance compared to negative past experience. These findings support **Hypothesis 2**.

With regard to the relationship between trust and reliance, one interesting result is that trust and reliance in agent teammates are not always correlated. In other words, higher trust levels do not always result in higher reliance. For instance, contrary to initial trust, initial reliance on agent teammate (effort levels in the first interaction) did not significantly differ between positive and negative experiences (see Figures 1(a) and 2(a)). We expected the initial effort levels 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. In the same way, no past experience condition lead to slightly lower reliance in the agent teammate (see Figure 3(a)), albeit significantly higher trust levels (see Figure 1(b)), compared to positive past experience. Similarly, when no past experience is compared with negative past experience, significant differences in reliance (see Figure 3(b)) are found despite similar trust levels (see Figure 1(c)). These somewhat counter-intuitive findings suggest that there exist factors affecting reliance other than trust. Lee and See [30] explain this phenomenon as "Trust guides –but does not completely determine– reliance". These results are in agreement with those obtained by previous studies [26].

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. The cumulative game results also suggest that greater reliance on agent teammate lead to less redundant work (see Tables 2 and 3) as expected.

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 led to significantly higher initial trust compared to negative past experience. However, the impact of past experience diminished with the increase in experience with the present teammate. On the other hand, no past experience led to significantly higher trust levels that lasted over interactions compared to positive past experience. Conversely no significant difference in trust levels is observed between no and negative past experience conditions. The effects of past experience are not limited to trust attitude. Positive (negative) past experience prompted (repelled) participants to rely on agent teammates towards achieving team goals. Lack of past experience, on the other hand, neither improved reliance like positive past experience nor worsened it like negative past experience. Lack of past experience, on the other hand, neither improves reliance like positive past experience nor worsens it like negative past experience. As a result, positive past experience engenders an increase in reliance on agent teammates and concomitantly a reduction in redundant work, which improves the team efficiency.

Operational lessons can be drawn from this study for agents to consider while making interaction decisions. Provided reasonable amount of information on past experiences of potential human teammate is available, agents should carefully consider the likely effect of presence and type of those experiences including relative ease of trust formation due to (a) positive feelings arising from positive past experience, (b) lower expectations as a result of negative past experiences, and (c) trust biases of humans without past experience. On the contrary, agents should be wary of situations where the human partner(s) may not trust the agent over interactions due to (a) high expectations formed from positive past experience, (b) negative feelings formed from negative past experience.

Hence, it is critical for agents to estimate the trust levels given the past experiences of potential human teammates.

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 (AAAI11)*. 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, 77–88.
[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] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies* 58, 6 (2003), 697–718.
[17] 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.
[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] Feyza Merve Hafizoglu and Sandip Sen. 2015. Evaluating Trust Levels in Human-agent Teamwork in Virtual Environments. In *HAIDM Workshop at the Autonomous Agents and Multiagent Systems (AAMAS'15)*. 1–16.

[23] 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.

[24] 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)*.

[25] Sebastian Hergeth, Lutz Lorenz, and Josef F. Krems. 2017. Prior Familiarization With Takeover Requests Affects Drivers' Takeover Performance and Automation Trust. *Human Factors* 59, 3 (May 2017), 457–470.

[26] 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.

[27] 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).

[28] 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.

[29] 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.

[30] John D. Lee and Katrina A. See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46, 1 (Spring 2004), 50–80.

[31] Jr. Lount, R. B. 2010. The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of Personality and Social Psychology* 98, 3 (2010), 420–433.

[32] 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.

[33] 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.

[34] Daniel J. McAllister. 1995. Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. *Academy of Management Journal* 38, 1 (February 1995), 24–59.

[35] Stephanie M. Merritt, Heather Heimbaugh, Jennifer LaChapell, and Deborah Lee. 2013. I Trust It, but I Dont 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.

[36] 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.

[37] Kathleen L. Mosier, Linda J.Skitka, Susan Heers, and Mark Burdick. 1998. Automation Bias: Decision Making and Performance in High-Tech Cockpits. *The International Journal of Aviation Psychology* 8, 1 (1998), 47–63.

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] Juergen Sauer, Alain Chavaillaz, and David Wastell. 2016. Experience of automation failures in training: effects on trust, automation bias, complacency and performance. *Ergonomics* 59, 6 (May 2016), 767–780.

[43] 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.

[44] 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.

[45] Piotr Sztompka. 1999. *Trust: A Sociological Theory*. Cambridge University Press, Cambridge, UK.

[46] 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.

[47] 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.

[48] 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.

[49] 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)*.

[50] Robert A. Wagner and Michael J. Fischer. 1974. The String-to-String Correction Problem. *J. ACM* 21, 1 (January 1974), 168–173.

Received November 2018; revised March 2019; accepted April 2019