

RESEARCH

Open Access



Trust transfer between contexts

Cody Buntain* and Jennifer Golbeck

*Correspondence:
cbuntain@cs.umd.edu
Department of Computer Science,
University of Maryland, A.V. Williams
Building, College Park, MD, USA

Abstract

This paper explores whether trust, developed in one context, transfers into another, distinct context and, if so, attempts to quantify the influence this prior trust exerts. Specifically, we investigate the effects of artificially stimulated prior trust as it transfers across disparate contexts and whether this prior trust can compensate for negative objective information. To study such incidents, we leveraged Berg's investment game to stimulate varying degrees of trust between a human and a set of automated agents. We then observed how trust in these agents transferred to a new game by observing teammate selection in a modified, four-player extension of the well-known board game, Battleship. Following this initial experiment, we included new information regarding agent proficiency in the Battleship game during teammate selection to see how prior trust and new objective information interact. Deploying these experiments on Amazon's Mechanical Turk platform further allowed us to study these phenomena across a broad range of participants. Our results demonstrate trust does transfer across disparate contexts and this inter-contextual trust transfer exerts a stronger influence over human behavior than objective performance data. That is, humans show a strong tendency to select teammates based on their prior experiences with each teammate, and proficiency information in the new context seems to matter only when the differences in prior trust between potential teammates are small.

Keywords: Trust transfer; Trust contexts; Trusting behavior; Investment game

Introduction

The trust one person has in another varies depending on context [1, 2]. You may trust someone to water your plants while you are away but not trust that person to babysit your children. This contextualization of trust makes sense; by definition, trust requires the truster to make herself vulnerable, taking a risk based on her belief that the trustee will act in her best interest. If the truster has good information about the trustee's plant tending skills, she may be willing to risk her greenery, but a green thumb does not imply childcare skills, so risking her children's safety does not follow.

Regardless of these differences, we often rely on trust "transferring" from one context to another. If our truster has a reliable co-worker, she may ask the co-worker to tend to her plants without any information about the trustee's horticultural skills. This transfer seems to occur despite contextual distinction because the truster's prior experience with the trustee's reliability increases her belief the trustee will behave appropriately even though she has no information about past performance in the new context. These apparent contradictions between observations of high trusting behavior and expectations of low trust are consistent with a plethora of prior research on the paradoxes of initial trust [3–6].

While the body of work on trust, how it evolves, dissolves, is stimulated, and how it can be transferred from entity to entity is relatively large, trust as it transfers *between contexts* has received little attention. Despite this lack of attention, however, inter-contextual trust seems to have noticeable effects on trust decisions. Indeed, we often see contextual trust transfer exploited in exercises designed to enhance trust between colleagues. Consider the trust fall or other team building exercises popular in corporate environments. These exercises are designed to build interpersonal trust that will then transfer to the workplace, even when the context of the exercise is largely unrelated.

Trust transfer is defined as “when a [truster] bases initial trust in [a trustee] on trust in some other related entity, or on a context other than the one in which the [trustee] is encountered,” and it is the latter half of this definition on which we concentrate [7]. Specifically, we are interested in how strongly trust transfers between largely unrelated contexts *and* how important that trust is in the face of other objective information. To test these concepts, we constructed a two-phase experiment and ran it with human participants recruited from Amazon’s Mechanical Turk platform. Each participant played Berg’s investment game, a standard construct often used to establish and test trust, against three other “players” controlled by agents programmed to stimulate varying degrees of trust [6]. Participants were then asked to pick one of the three agents from the previous game as a teammate for a partner-based cooperative game based on Milton Bradley’s *Battleship* that included the possibility of betrayal. We found that not only did participants choose the most trustworthy agent as their teammate the vast majority of the time, but they chose the trustworthy agent even when objective information about its skill at the *Battleship* game indicated the agent was less qualified (sometimes significantly so) than other possible choices.

We believe these experiments are an interesting empirical analysis of trust transfer across contexts and support quantifying the strength of prior trust compared to objective information.

Background and related work

The idea that a person’s trust in another person or the social structures in which they interact influences behavior and performance is not new. As early as the 1920s, Edward Thorndike referred to a cognitive bias in which one’s initial impressions of another’s character exerts significant influence over later subjective assessments of that person as the “Halo Effect” [8]. Given the potential social benefits of such increased trust, a significant subfield of research has grown around topics like techniques for studying trust, stimulating trust in new relationships, repairing trust after betrayal, and transferring trust from one entity to another [1, 3–5, 9]. Mechanisms for transferring trust are of particular interest because they facilitate many daily interactions like shopping at a new location of a known brand or visiting a restaurant recommended by a friend. While trust transfer between contexts is not well understood, existing works on trust transfer, discussed below, provide an essential foundation for our work. Similarly, since trust can be both difficult and expensive to study in situ, researchers have developed constructs for studying trust using controlled game environments, a technique we leverage, so we include a brief background on this game research below as well.

Studying trust with games

In the mid-nineties, Berg, Dickhaut, and McCabe introduced a game construct to study trust and reciprocity with a simple two-player investment game [6]. This investment game has been a popular lens through which researchers have explored methods to support or suppress trust between players. Berg brought together a number of undergraduate students to play a double-blind, one-shot game in which players were randomly split into two groups: one of investors and one of investees (Berg et al. used “trustors” and “trustees”). All players were provided an initial sum of money for participation, and investors were then instructed to choose some amount of this sum to send to, or “invest with,” the investee. Investees received triple the investor’s investment and chose some portion of this sum (or none at all) to return to the investor. Investors were said to trust the investees if they invested any amount as there was no guarantee any amount would be returned, and investees were said to reciprocate trust if the amount returned was greater than the investor’s investment.

Game theoretic literature provides a solution to Berg’s game in the form of a subgame perfect equilibrium in which optimal play is to invest nothing and return nothing. Berg’s results, however, showed a non-zero average investment and return, which ran counter to the predicted equilibrium. From these results, Berg, Dickhaut, and McCabe concluded that self-interest could not account for these deviations from the expected subgame perfect actions. Instead, they theorized that some investee players returned larger paybacks because they were entrusted with larger investments; that is, they reciprocated the trust placed in them by the investors. Since then, this simple game construct has served as a proxy for evaluating trust and reciprocity, giving researchers a tool for quantifying how changes in information and behavior can affect trust.

One important modification to Berg’s investment game has been the addition of multiple rounds of play. In 2004, Cochard et al. investigated this extension by comparing a seven-round repeated game with Berg’s one-shot version [10]. From this work, Cochard theorized investor behavior was consistent with a “reciprocity hypothesis” in which investors would trust more and therefore invest more as investees reciprocated more. Engle-Warnick and Slonim also investigated such multi-round modifications and determined that, in certain scenarios, investor trust and reciprocity could be maintained or strengthened as rounds increased [11]. Additionally, they observed that investor trust resets to the player’s default level (governed by an external propensity to trust) when paired with a new partner.

We base our implementation of Berg’s investment game on these latter results and use it as a mechanism to foster trust between parties. In this manner, we can leverage the amount of reciprocity as a *proxy* for trust; that is, the more reciprocation in the investment game, the higher the trust between the players.

Trust transfer

Trust transfer, or the transitivity of trust, has been studied in a variety of ways and applied across many disparate environments, from team formation to e-commerce to distributed computer systems. While Stewart’s definition allows trust to transfer from an entity/context to a separate entity/context, most treatments concentrate only on transferring trust to a new, separate, and distinct entity (generally in the same context) [7].

Trust transfer from entity to entity is the more commonly studied occurrence and one on which the majority of recommendation systems operate. Though not often explicitly identified, recommendation systems rely on the trust a user has in the recommender (either a friend, or website) to *transfer* to the entity being recommended. This effect was shown in Strub and Priest's 1976 discussion on trust transfer as an "extension pattern" for drug users to expand the social networks necessary to procure desired drugs using a trust third party as a referer and mediator to new entities [12]. Stewart presents other examples of trust transfer as well (see [7]). More recently, research into trust transfer has migrated from psychology and sociology to computational fields as trust in distributed networks, social networks, and e-commerce sites becomes increasingly important. The role of trust transfer in these new areas is unclear, however, as seen in Christianson and Harbison's 1997 work on why trust transitivity is dangerous in security protocols [13]. Golbeck's 2005 work gives a brief overview on other treatments where trust transfer occurs and propagates on the World Wide Web; she then goes on to describe an algorithm, TidalTrust, for using trust transitivity to propagate trust values across social networks for friends of friends and movie reviews [14]. Dong, Russello, and Dulay's 2007 work investigated the constraints necessary for trust to transfer between actors in such distributed, decentralized networks [15]. Their work presented a simple model and set of logical formulae for formally describing trust transfer based on trust policies. Furthermore, in works by Stewart and by Levin and Cross, researchers studied how weak ties enabled trust transfer between communities with specific interest in hyperlinks among websites, which support a form of trust transfer from a trusted website to another site [7, 16].

Regarding trust transfer from context to entity, McKnight, Cummings, and Chervany investigated the mechanisms for initial trust formation in new relationships (i.e., relationships where the actors have no prior experience with each other). While their work was not directly about trust transfer, they posited one's context was an important hidden factor in explaining unexpectedly high levels of initial trust because certain "institutional cues that enable one person to trust another without firsthand knowledge" could enforce social norms or positive behaviors [5]. Riegelsberger, Sasse, and McCarthy expanded on contextual factors that incentivize trust by including temporal and social properties along with aforementioned institutional cues in a new trust model [9]. Social context influence includes factors like social norms and reputation, and temporal context covers factors like length of interaction or whether the interaction will be repeated. These contexts make intuitive sense: interactions with a potentially threatening individual likely differ based on where those interactions occur (e.g., a crowded room versus a dark alley). Likewise, less incentive exists to behave in a trustworthy manner if two entities will only interact one time and then never meet again.

Inter-context trust transfer for the same entity seems to be both a relatively new and rarely studied concept. Though cross-context transfer occurs regularly in our daily lives, the main treatment of it as a research question seems to be the recent work on how consumers transfer their trust in retailers' brick-and-mortar stores to their online stores. Stewart introduced this retailer cross-context trust in 1999 and followed it up in 2003 with an experiment that showed users' trust intentions and intentions to buy from an online retailer were increased if the retailer's site showed a picture of a building or physical retail location [7]. Riegelsberger and Sasse demonstrated similar results in a 2001 study on cues

online retailers could employ to increase user trust in their sites: specifically, “Inexperienced e-shoppers are likely to transfer trust: They will give on-line shopping a first try with retail organizations they are familiar with” [17]. This cross-context trust transfer has also evolved as web commerce has grown in popularity and complexity. In 2011, Lu et al. published a paper on consumer trust transfer from web-based payment services to mobile-based services in much the same manner consumers transferred trust from brick-and-mortar stores to web sites [18].

While these investigations into the flow of user trust from physical buildings to Internet locations to mobile services do not exactly mirror our experiments on trust across game contexts, strong parallels exist that support our work. For example, in these related works, the user’s objective to purchase an item or service is consistent, and the details around how the purchase is made changes from physical to web to mobile contexts. Similarly, in our implementations of both Berg’s investment game and the modified Battleship game, the user’s objective of playing a game to accumulate money is consistent, but the contextual details around how the games are played and the structural rules in place to limit betrayal change. Furthermore, our work takes this exploration a step further as we have no knowledge of any other investigation into the strength of cross-context trust with respect to new objective information.

Methodology

Our goals in the experiments outlined in this paper were two-fold: 1) to confirm whether artificially stimulated trust between players in one context would transfer to an unrelated context; and 2) whether such experiential prior trust would offset or overcome quantitative player information. To this end, we constructed a two-phase experiment in which the first phase established varying degrees of trust between a human player and three automated agents in Berg’s investment game. The second phase then asked the human participant to select one of these three agents as a teammate to play in a separate, unrelated game. We grouped our experiments into two treatments based on the information provided to the human participant at the beginning of the second phase of play. In the first treatment, participants were shown data on the investment and return amounts from the previous rounds. As mentioned, this investment and return data denotes the amount of reciprocity in the game, which we use as a proxy for trust (i.e., higher investment and returns implies higher reciprocity and higher degrees of trust).

In the second treatment, participants were provided with this reciprocity data in addition to information regarding the *proficiencies* of each agent. Proficiencies here took the form of objective information on the percentage of prior second-phase games that particular agent had won. By varying the pairs of reciprocity and proficiency scores, we could then determine how strong the influence of prior trust was in teammate selection. Statistical analysis of these teammate choices would then hopefully inform us about the effects of trust transfer across contexts.

Game constructs

The experiments described herein leverage two different game constructs: Berg’s investment game, and a customized, cooperative, two-versus-two version of Milton Bradley’s Battleship. While our use of Berg’s game is fairly standard with an initial investor sum of \$10 and investment multiplier of three, the three agent strategies used were non-standard

and were designed to elicit a specific response from the human player. This strategy set consisted of the following:

- **Greedy Agent** Never return any investment (subgame perfect strategy)
- **Benevolent Agent** Always return slightly more than the investor invested
- **Exploitive Agent** Return slightly more than invested in first two rounds, then return nothing

Since the goal in deploying these strategies is to stimulate varying degrees of trust between the human participant and the agents, we ascribe an ordering to these agents based on the strength of trust they should instill. Our definition of “trust” follows that of Stewart: a trustworthy agent’s behavior should be “benevolent, competent, honest, predictable in a situation” [7]. Therefore, we expect the benevolent agent to be the most trustworthy and exploitive agent to be the least trustworthy with the greedy agent falling somewhere in between.

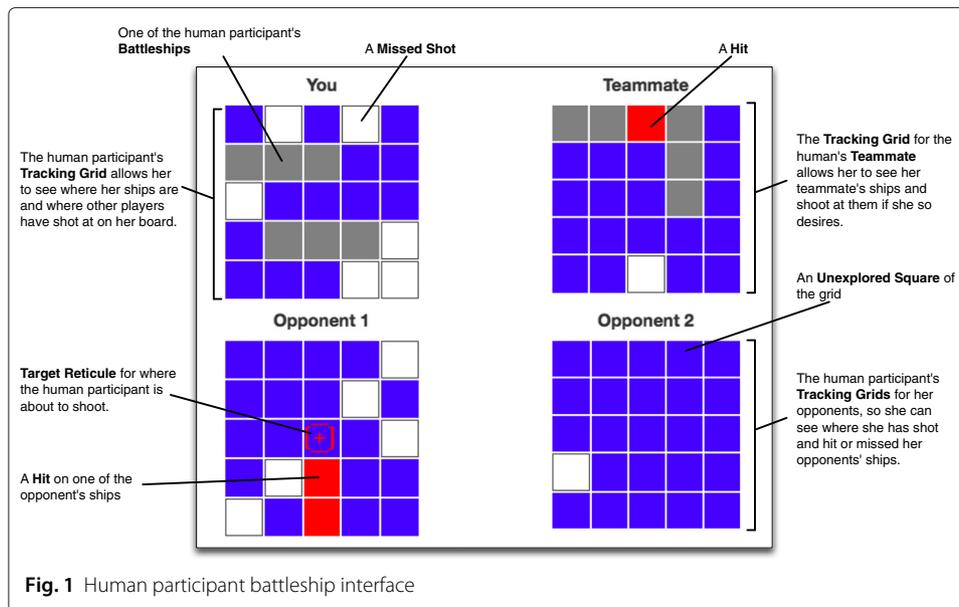
While our estimate of the exploitive agent’s trustworthiness may seem counterintuitive, we refer to the Lewicki and Stevenson’s 1997 paper that presents a model of trust across three different axes (calculus-based, knowledge-based, and identification-based trust). Their definition of knowledge-based trust supports the greedy agent’s trustworthiness being higher than the exploitive agent’s by virtue of predictability. Since the greedy agent always behaves consistently, it is more predictable than the exploitive agent, and “predictability enhances trust - even if the other is predictably untrustworthy - because we can predict the ways that the other will violate the trust” [4].

Our modified, four-player Battleship game was slightly more complicated than our implementation of Berg’s game. The original version of Milton Bradley’s Battleship is a two-player game in which each player has a fixed number of ships of known sizes. Each player places her ships on a private, hidden 10×10 board and tries to locate all her opponent’s ships on a separate 10×10 tracking board. In our version, each player is given two boats, each three consecutive squares long, that are placed on the her 5×5 board randomly. Each player also is presented with three additional 5×5 tracking board, one for each other player, so the player can track where her ships are, her teammate’s ships are, and where she has searched for her opponents’ ships.

After placement, each player then selects a square from one of the four 5×5 boards simultaneously and “shoots at” that square to see whether the opponent has a ship at that location or not. The objective of the game is for a player and her teammate to find all their opponents’ ships and “sink” them by shooting all squares of each opponent ship before the opponents sinks her and her teammate’s ships. Figure 1 illustrates how these boards look to the human participant.

Teams are divided into two sets of two players, and players on the same team can see locations of each other’s boats. Critically, each player can shoot at any board regardless of team affiliation (even her own). Also, rather than players’ taking turns targeting opponent ships, in our versions, shots are taken simultaneously; that is, all players select their targets at the same time, shots are revealed simultaneously, and shot sources are kept hidden, so no player knows who shot at which target or which board.

Play ends when all the boats on one team have been sunk, and winnings are divided up among the surviving players on the winning team. This division of rewards provides



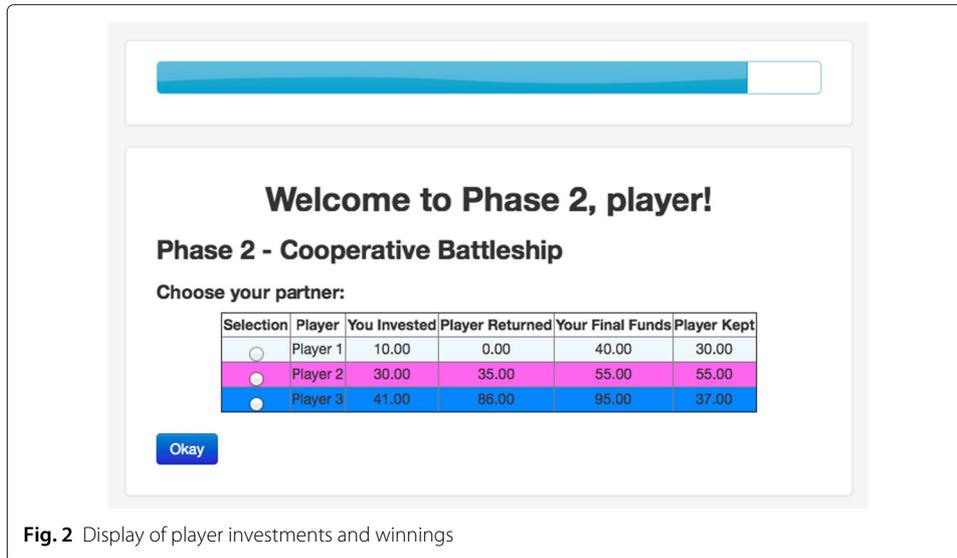
some incentive for betrayal if a player wants to kill her teammate and keep all winnings for herself. The dynamics of this Battleship game are interesting in that players do not necessarily know if/when their teammates have betrayed them and must therefore place some level of trust in their teammates during play.

Experimental design

Our experiment was made available through a web interface in which participants were presented with instructions on how the game would be played and a short quiz to ensure appropriate understanding of the game. After reading the instructions and successfully completing the quiz, the participant was randomly assigned to one of the game treatments before starting the first phase of the experiment. In this first phase, the participant would play five (5) consecutive rounds of Berg's investment game with each of three automated agents (participants were not aware they were playing automated opponents).

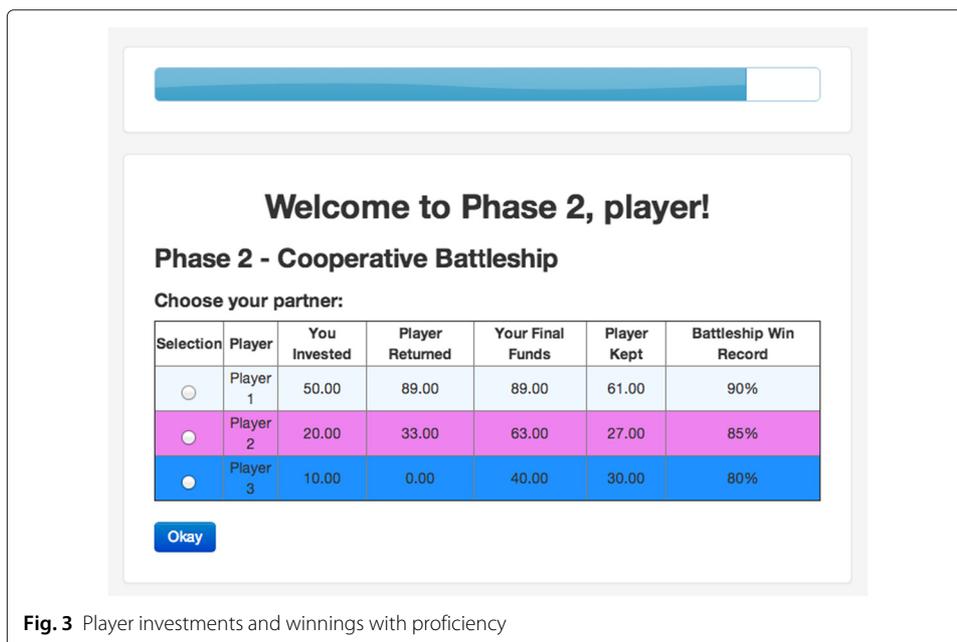
In each Phase 1 round, the participant was given a sum of \$10 that she could keep or invest with the current investee agent. After choosing the investment amount, that amount was multiplied by the investment factor and sent to the agent to be processed by its strategy. The participant was then shown how much the agent returned before moving on to the next round where she would be provided with a new sum of \$10 for investment. At the end of each set of five rounds, the participant's total investment, winnings, and remaining funds were calculated and stored. It is worth noting here that the investee agents were color-coded such that, after five rounds of play with one agent, the background color of the web page would change to make the transition to a new investee more explicit to the participant. Background colors were chosen randomly to avoid any color-related emotional bias.

Once the participant completed each set of games with the three agents, she was then shown the instructions and rules for the battleship phase of the experiment. After reading these rules, she was then provided with a tabular breakdown of her investments, her winnings, investee returns, and investee winnings (see Fig. 2) and asked to select one of these players as her partner for the next phase.



Depending on the game treatment being played, the participant might also be shown an additional column that lists each player’s “Battleship Win Record” (Fig. 3), which represented the percentage of prior Battleship games won by that agent.

Since the choice of teammate in the second phase was the main point of interest, actual play in the Battleship game was not monitored. The subject simply played with three automated computer players who fired randomly. When the Battleship game ended, however, we did record whether the participant’s team won and whether the participant’s teammate survived as this information was necessary for assigning end-of-game bonuses (discussed below).



Game treatments

As mentioned, these experiments included two different treatments. In the first treatment, human participants were not provided any information regarding agent proficiencies in the Battleship game; that is, the only information provided was the data on investments and returns. In the second treatment, each agent was assigned a win-rate proficiency on a 0–100 scale, and that information was displayed to the participant prior to teammate selection. This second treatment consisted of cases based on the relative differences between these proficiency values. Assignments of agent strategy to proficiency in this second treatment were random. The main motivation for the cases here was to identify if a limit existed on how much of a difference in proficiency prior trust could mitigate. These cases are shown in Table 1.

Participant selection

To obtain a large and diverse data set in a cost-effective manner, we leveraged Amazon’s Mechanical Turk framework to recruit participants via the Internet. These players were paid \$0.95 for participating in the experiment and were given a bonus based on their performances in both phases to incentivize good play. While first-phase rounds had clear dollar amounts at stake, to provide a similar incentive for the Battleship game in the second phase, participants were told the winning team would get \$50 to be divided among surviving team members. Bonuses were then calculated by taking the number of dollars the participant had at the end of the first phase and adding the winnings from the Battleship game (either \$25 or \$50 if the participant’s teammate did not survive) and paying a penny for every dollar won.

To participate in the experiment, each human player accepted a Human Intelligence Task (HIT) from the Mechanical Turk interface, which allowed us to enforce a number of qualifications prior to play. First, players were only allowed to take part in the experiment once; that is, after completing both game phases, participants were not allowed to play the game again. Second, all players were required to have a high approval rating (greater than 96 %) from other Mechanical Turk requesters in order to take part. Third, all players must have completed more than 60 HITs in Mechanical Turk. Lastly, players were required to be in the United States to minimize confounds from cultural differences.

Results

The following section demonstrates the effects of prior trust across our treatments and 684 unique human participants. All results are reported using a chi-squared test of independence between the agent type and whether it was selected as the teammate for the second phase. As our experiments for the second treatment were constructed such that proficiency values were distributed equally among agent types, expected selection counts were nearly uniform. In short, the results below show a significant relation between agent

Table 1 Treatment #2 per-case proficiency values

Case	Proficiencies
1	[80, 85, 90]
2	[67, 80, 91]
3	[20 ± 7, 52, 85 ± 7]
4	[3 ± 2, 52, 85 ± 2]

type and teammate selection regardless of proficiency information, or more plainly, participants were much more likely to select the most trustworthy agent (the benevolent agent) across all treatments.

Treatment #1: No proficiency information

For this first treatment, participants played the three sets of Berg’s investment game and then selected their Battleship teammates based solely on prior experience. In total, we collected data from 175 participants in Mechanical Turk. As expected, participants overwhelmingly chose the trustworthy benevolent agent as the teammate when no additional information is available. Data on this activity is shown in Table 2 with mean phase-one winnings \bar{W}_1 and teammate selection counts C . If teammate selection was independent of the trust established in the investment games, one would expect per-agent-type selection counts to be uniform, or that each agent would be selected as a teammate approximately 58 times, which is demonstrably false. The chi-squared test for independence supports this strong relation between agent type and teammate selection with $\chi^2(2, n = 175) = 399.00, p < 0.001$.

Treatment #2, Case 1: High proficiency, small differences

In the second treatment, participants selected their Battleship teammate based on prior experience and objective proficiency data. For this first case, participants were shown high-proficiency values with small differences between them ($\Delta = 5$). From the 208 participants collected for this treatment, users again significantly preferred the trustworthy agent over any other agent regardless of proficiency values. Table 3 illustrates these selections with mean phase-one winnings \bar{W}_1 , mean proficiencies $\bar{\rho}$, teammate selection counts C .

As in the first treatment, one would expect teammate selections to be uniformly distributed across agent types if proficiency was the dominant factor here, and once again, this uniformity is not present. Our data and the chi-squared test for independence instead supports the significant relation between agent type and teammate selection with $\chi^2(2, n = 208) = 376.20, p < 0.001$.

Treatment #2, Case 2: High proficiency, large differences

This case followed a similar setup as the first case of Treatment #2 except with larger differences in proficiency ($\Delta \approx 12$). Surprisingly, the 182 users who participated in this treatment *still* preferred the trustworthy agent over the other two agents, even with the larger differences in proficiency. In fact, even when the most trustworthy agent was listed as the objectively *worst* teammate, participants still chose the trustworthy agent the vast majority of the time (44 times out of 56). For completeness, Table 4 tabulates these

Table 2 Mean winnings \bar{W}_1 and teammate selection count C without proficiency

Agent type	\bar{W}_1	C
Benevolent	80	160
Exploitative	44	10
Greedy	38	5
Total		175

Table 3 Case 2.1 – Mean winnings \bar{W}_1 , mean proficiency $\bar{\rho}$, and selection count C

Agent type	\bar{W}_1	$\bar{\rho}$	C
Benevolent	80	85	177
Exploitative	43	85	15
Greedy	37	85	16
Total			208

behaviors. As was the case in the previous treatments, the chi-squared test for independence again shows a significant relation between agent type and teammate selection with $\chi^2(2, n = 182) = 388.08, p < 0.001$.

Treatment #2, Cases 3 and 4: Very low proficiency, large differences

These last cases explored whether participants would still choose the agent with which they developed the most “trust” even if that agent’s proficiency was significantly worse than the other options. As in the prior cases, participants overwhelmingly selected trustworthy agents over the others, seemingly without regard for proficiency. Tables 5 and 6 show tabulated versions of these results.

Overall importance of trust and proficiency

If participants were choosing based only on trust, they would always choose the benevolent agent as their Battleship teammate. If participants choose based on proficiency, they should always choose the most proficient agent. To understand the interplay, we looked at how often the benevolent agent was chosen when it was the most proficient player and not the most proficient player, and how often non-benevolent players were chosen when they were the most proficient or not the most proficient. Figure 4 shows these rates.

We can see here that the benevolent agent is chosen the vast majority of the time, regardless of whether it is most proficient or not. This is a strong indicator that trust is far more important to subjects than the objective information about skill.

There is another interesting data point here. There are cases where subjects chose a partner who was both non-benevolent *and* non-proficient. Why would a subject ever choose a partner she did not trust and who was objectively bad at the game? To find out, we sent followup questions to subjects in this group asking them why they chose the partners they did. The responses indicate other “non-rational” factors are at work:

I choose the one guy who returned \$0.00 to me every single time. I knew he was out for himself; having him on my “team” meant having my “worst enemy” in a position where he wasn’t directly “against” me, and I’d know what he was up to.

-Subject 31

Table 4 Case 2.2 – Mean winnings \bar{W}_1 , mean proficiency $\bar{\rho}$, and selection count C

Agent type	\bar{W}_1	$\bar{\rho}$	C
Benevolent	81	79	155
Exploitative	43	79	18
Greedy	38	80	9
Total			182

Table 5 Case 2.3 – Mean winnings \bar{W}_1 , mean proficiency $\bar{\rho}$, and selection count C

Agent type	\bar{W}_1	$\bar{\rho}$	C
Benevolent	81	49	35
Exploitative	43	51	8
Greedy	38	51	4
Total			47

The first thing I did was attack the crap out of my teammate. I knew where his ships were, and I blew him out of the H2O.

-Subject 54

While Subject 31 may not have thought deeply about his strategy (i.e. having that “worst enemy” on his team would not prevent him from acting badly, nor did we say that unchosen players would be on the opposing team), he clearly had identified this agent as “bad” and made a conscious choice to partner with it because of that. Subject 54 has a much clearer intention: revenge. He was willing to give up potential monetary rewards for the satisfaction of punishing the player who treated him poorly.

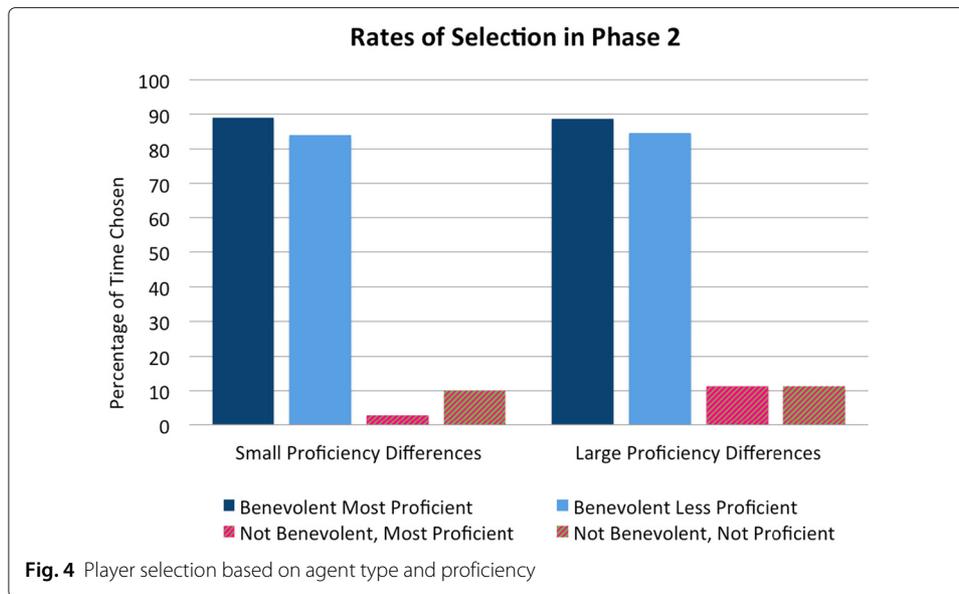
Along this same vein, we also compared the relative importance of trust versus proficiency by analyzing the differences in return (as a proxy for trust) and differences in proficiency between agents selected for teammates and those remaining. For each game, we generated four data points: the differences between the selected agent and the two remaining agents and their negatives. That is, if Agent x was selected as the teammate, and Agents y and z were remaining, we would generate the points $(x.return - k.return, x.proficiency - k.proficiency)$ and $(k.return - x.return, k.proficiency - x.proficiency)$ for $k \in \{y, z\}$. Each point was labeled based on whether the agent on the left of the subtraction was the selected teammate. We then fed these labeled points into a linear support vector machine to determine the relative weights of trust versus proficiency (shown in Table 7) and characterize the decision plane separating the chosen agents from the rest. Figure 5 shows this decision plane with green dots indicating points for selected agents and red for unselected agents. From these results, it is clear that return, as a proxy for trust, is significantly more influential in deciding teammate selection than proficiency, even when the trusted teammate’s proficiency is near zero. Another interpretation of these results is that differences in proficiency only become important factors in teammate selection when the differences between returns is near zero.

Discussion

To restate, this research had two main objectives: confirm that establishing trust in one context would transfer to and influence behavior in an unrelated context, and identify to what extent prior trust could compensate for a poor objective evaluation of that subject. While intuition and factors like the “Halo Effect” seem to confirm the answer to the first

Table 6 Case 2.4 – Mean winnings \bar{W}_1 , mean proficiency $\bar{\rho}$, and selection count C

Agent type	\bar{W}_1	$\bar{\rho}$	C
Benevolent	80	47	48
Exploitative	44	45	6
Greedy	38	46	3
Total			57



question, our statistical results from the first empirical treatment further concretize this hypothesis. Furthermore, results for the second and third treatments strongly indicate that human players value prior experiential trust over quantitative information regarding potential teammates. At least for the context of games with small monetary rewards, trust established in an earlier context clearly influences behavior even when the two contexts differ, and non-trust-related quantitative information only seems important when differences in this trust are small.

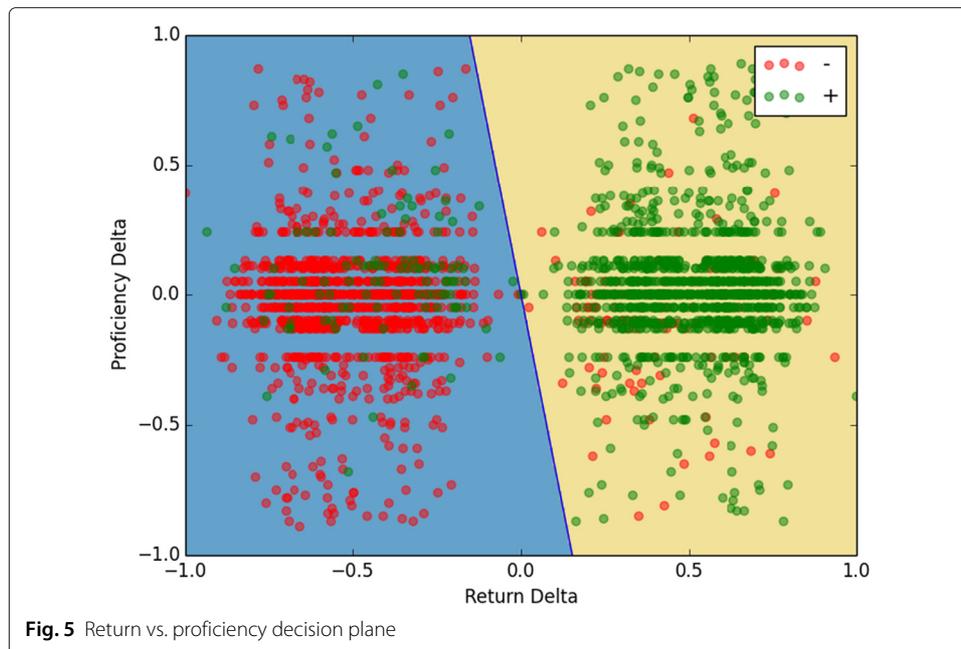
Limitations and future work

As with many experiments involving human participants, it is difficult and expensive to get large amounts of data, and many confounding factors might be present. The chief potential confound present in this work is difficulty separating the two contexts discussed. That is, Berg’s investment game and the modified Battleship game we employed may not be significantly different enough for a human participant to truly see them as two separate contexts. Though the rules of the two games are quite different, and the teammate in the Battleship game has much more power to betray the human participant than in the investment game (e.g., the participant can choose to stop investing money with an agent but cannot choose to stop being an agent’s teammate), in a fundamental way, both contexts are simple, faceless games for money on the Internet. Though the existing literature supports our claims, the potentially small contextual differences may skew our ability to quantify the influence of prior trust [7, 17, 18].

It may be possible to address this issue in future work by replacing the Battleship game with a significantly different approach completely outside the realm of online games.

Table 7 Weights for return vs. proficiency

Dimension	Weight
Return	3.1
Proficiency	0.47



One such scenario might be to ask the human participant whether she would be willing to add one of the agents on a popular social network (e.g., as a friend on Facebook or follower on Twitter). The social networking context is clearly much different than this gaming context, but it then becomes difficult to address how one might integrate objective information into such an experiment.

Implications

These results have several interesting implications across a number of areas, likely foremost of which is the generalization of human preference to follow prior experience over quantitative information. This preference could be both helpful and harmful; one could leverage simple trust-based tasks to bootstrap trust in automated systems or collaboration with new partners. At the same time, however, if one has a poor early experience with a system or other user, even if it was in a different context, this poor experience will likely affect further use regardless of any informational assurances of improvement.

Recommendation system designers might also benefit from this trust transfer effect across contexts. For example, if a user *A* follows the recommendations of another user *B* in one context (movie recommendations for example), that implicit trust between users could be transferred to other recommendation system contexts (e.g., book or restaurant recommendations). Given our results, there is reason to believe that, if user *A* is told user *B* also recommended some set of books or restaurants, user *A* is more likely to follow those recommendations over those of another user *C* with whom user *A* has no experience, even if user *A* is told user *C* completely matches user *A*'s preferences.

Another surprising result of this work is the robustness of transferred trust to critical quantitative data. As mentioned, in the case with large differences in reported proficiency, the difference between the best and worst values is $91 - 67 = 24$, and yet the majority of participants still selected the most trustworthy agent as the teammate even when

these agents had a very low performance indicators. These results are consistent with the existing literature regarding people's ability to dismiss information counter to their beliefs [5].

In a more abstract sense, these results could indicate the existence of a latent hierarchy in which trust can transfer horizontally across contexts but not necessarily up the hierarchy to other classes. While our results demonstrate trust does transfer well across monetary-game contexts even if the low-level implementations and purposes of those games vary widely, one might still expect trust established in this game context would not transfer to the context of taking care of one's children. As such, an interesting question raised by this work concerns the structure of this hierarchy; what are the few modalities of trust that are closest to the root of this hierarchy, and how can we better understand the mechanisms through which trust transfers across this hierarchy? Integrating existing research from the social sciences may facilitate this research.

Conclusions

We sought to confirm whether the trust transfer effect is observable across different contexts and to what extent this trust transfer can compensate for objective information on agent performance. Our results demonstrate this trust transfer effect does in fact exist across disparate contexts and greatly influences participant behavior even when provided quantitative data contradicts this trust. As listed, this work has implications for several different fields and provides a solid foundational step for further exploration of how humans transfer trust between varying environments and contexts.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CB developed the source code for the game constructs and carried out the initial and follow-up experiments on Amazon's Mechanical Turk framework. JG provided funding for this study and to travel to the workshop at which the shorter form of this paper was published. Both authors contributed to the study's conception, performed analysis on the results, and helped draft the manuscript. Both author's read and approved the final manuscript.

Acknowledgments

Several people participated in the refinement and development of the game constructs used in this experiment, with special thanks to M. Leigh Cook and Phillip Dasler for game testing and manuscript review. Additionally, thanks to the workshop co-chairs and program committee for the AAAI 2014 Workshop on Incentives and Trust in E-Communities (WIT-EC'14), where the origins of this work were first published. This work was partially supported by the Army Research Laboratory and the Network Science Collaborative Technology Alliance.

Received: 11 December 2014 Accepted: 25 May 2015

Published online: 12 June 2015

References

1. Deutsch M (1977) *The Resolution of Conflict: Constructive and Destructive Processes*. 1st edn. Yale University Press, New Haven, CT
2. Shapiro SP (1987) The social control of impersonal trust. *Am J Sociol* 93(3):623–658. doi:10.1086/228791
3. Jones GR, George JM (1998) The experience and evolution of trust: Implications for cooperation and teamwork. *Acad Manag Rev* 23(3):531–546. doi:10.5465/AMR.1998.926625
4. Lewicki RJ, Stevenson MA (1997) Trust development in negotiation: Proposed actions and a research agenda. *J Bus Prof Ethics* 16(1):99. doi:10.2307/27801027
5. McKnight DH, Cummings LL, Chervany NL (1998) Initial trust formation in new organizational relationships. *Acad Manag Rev* 23(3):473–490
6. Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity, and social history. *Games Econ Behav* 10(1):122–142. doi:10.1006/game.1995.1027
7. Stewart KJ (2003) Trust transfer on the World Wide Web. *Organ Sci* 14(1):5–17. doi:10.1287/orsc.14.1.5.12810
8. Thorndike EL (1920) A constant error in psychological ratings. *J Appl Psychol* 4(1):25–29. doi:10.1037/h0071663

9. Riegelsberger J, Sasse MA, McCarthy JD (2005) The mechanics of trust: A framework for research and design. *Int J Hum Comput Stud* 62(3):381–422. doi:10.1016/j.ijhcs.2005.01.001
10. Cochard F, Van PN, Willinger M (2004) Trusting behavior in a repeated investment game. *J Econ Behav Organ* 55(1):31–44. doi:10.1016/j.jebo.2003.07.004
11. Engle-Warnick J, Slonim RL (2004) The evolution of strategies in a repeated trust game. *J Econ Behav Organ* 55(4):553–573. doi:10.1016/j.jebo.2003.11.008
12. Strub PJ, Priest TB (1976) Two patterns of establishing trust: the marijuana user. *Sociol Focus* 9(4):399–411
13. Christianson B, Harbison WS (1997) Why isn't trust transitive? In: Lomas M (ed). *Security Protocols*. Springer, New York, NY. pp 171–176. doi:10.1007/3-540-62494-5_16
14. Golbeck JA (2005) Computing and applying trust in web-based social networks. PhD thesis, University of Maryland, College Park. doi:10.1016/j.jpsychires.2008.02.003. <https://www.cs.umd.edu/~golbeck/pubs/Golbeck%20-%202005%20-%20Computing%20and%20Applying%20Trust%20in%20Web-based%20Social%20Networks.pdf>
15. Dong C, Russello G, Dulay N (2007) Trust transfer in distributed systems. In: Etalle S, Marsh S (eds). *Trust Management*. Springer, New York, NY Vol. 238. pp 17–29
16. Levin DZ, Cross R (2004) The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer. *Management Science* 50(11):1477–1490. doi:10.1287/mnsc.1030.0136
17. Riegelsberger J, Sasse MA (2001) Trustbuilders and trustbusters: the role of trust cues in interfaces to e-commerce applications. In: 1st IFIP Conference on E-Commerce, E-Business, E-Government, vol. 1. http://hornbeam.cs.ucl.ac.uk/hcs/publications/Riegelsberger+Sasse_Trustbuilders%20and%20trustbusters_I3E2001.pdf
18. Lu Y, Yang S, Chau PYK, Cao Y (2011) Dynamics between the trust transfer process and intention to use mobile payment services: A cross-environment perspective. *Information and Management* 48(8):393–403. doi:10.1016/j.im.2011.09.006

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
