Whom do you trust: The impact of facial emotion and behavior on decision-making

Timothy Ryan Campellone

A thesis submitted in partial satisfaction

of the requirements for the degree of

Master of Arts in Psychology

in the Graduate Division of the University of California, Berkeley

Committee in charge:

Professor Ann M. Kring, Chair

Professor Dacher Keltner

Professor Sheri Johnson

Spring 2012

Whom do you trust? The impact of facial emotion and behavior on decision-making

Consider the following scenario: you walk on to a used car lot and are immediately greeted by a smiling salesperson. A friend warns you that this particular salesperson has a history of withholding crucial information about the car in the interest of making a sale. Do you heed the advice of your friend and choose a different lot, or is the information signaled by the salesperson's smile enough to engender trust? Depending on what you use to inform your decision, you could end up being duped into overpaying for an inferior car. In this example, the salesman's smiling facial expression is incongruent with his behavioral intention of selling a car for more than it was worth. Decisions regarding whether a person is trustworthy, regardless of whether he or she is smiling, can be accurately made based on facial features within 100ms (Todorov, Pakrashi, & Oosterhof, 2009). However, the information we rely on to facilitate social decision-making is often in conflict, such as when facial displays of emotion are incongruent with behavior.

Researchers investigating social decision-making have utilized behavioral economic paradigms as a way to study human interactions while maximizing experimental control over variables of interest (for review, see Rilling & Sanfey, 2011). One paradigm in particular, the Trust Game (Berg, Dickhaut, & McCabe 1995), is well suited for investigating how people make decisions about whether to trust a social partner[1]. During the original Trust Game, two people were seated in different rooms. The person in the role of Player A was given $10 and the option to invest any amount of the $10 in Player B. The amount given to Player B was multiplied by the experimenter, and Player B was then instructed to return any amount of the multiplied sum to Player A. The amount invested by Player A represents how much that person trusts Player B to reciprocate the investment, with more invested money indicating greater trust. Player B's decision to return money to Player A reflects trustworthy behavior, with lesser returns indicating untrustworthy behavior. Diminished trust, therefore, is reflected by a small (or no) investment by Player A accompanied by a lack of reciprocation by Player B. The conservative or "safe" choice would be for Player A to keep the $10 and invest nothing. This way, Player A is guaranteed to leave the interaction with a positive gain. However, 30 of the 32 participants in the position of Player A gave some amount of money to Player B (average Player A investment was $5.16). Thus, without knowing anything about the other person with whom they were playing the game, most participants in the role of Player A were willing to invest money (i.e., trust the other player) in hopes that Player B would reciprocate their trust.

---

[1] In the original study conducted by Berg and colleagues, the game was called the Investment Game. However, since other studies employing this paradigm have referred to the Investment Game as the Trust Game, we will henceforth refer to the Investment Game as the Trust Game for the purposes of clarity.

Why do people decide to trust? Indeed, the decision to trust is a potentially risky choice. Participants' decision to invest an average of half the allotted money defies theoretically predicted behavior of a zero investment (Camerer, 2003). While previous studies have examined interactions between anonymous partners, more recent studies have looked at different social factors that influence decisions to trust, including the perceived trustworthiness of interaction partners. For example, participants in one study engaged in a series of single Trust Game interactions with different simulated partners represented by black and white emotionally neutral faces. Participants invested more in faces that were rated as trustworthy looking (van 't Wout & Sanfey, 2008). A similar study (Chang et al., 2010) used a repeated version of the Trust Game with four simulated players represented by black and white emotionally neutral faces that were selected based on independent and *a priori* trustworthy ratings. Simulated player trustworthiness influenced decisions to trust during the first interaction, but did not influence subsequent decisions. Instead, decisions during subsequent interactions were more influenced by the simulated player's behavior (i.e., whether the simulated player returned invested money). That is, players who reciprocated participant trust by returning a greater share of the invested sum were trusted more, even if that player's face was untrustworthy.

Taken together, these studies suggest that decisions to trust can be influenced by perceived trustworthiness signaled via facial features as well as by reciprocated investment. However, as the used car salesman example illustrates, facial displays of emotion are also potent signals during social interactions, communicating both feelings and intentions (e.g., Keltner & Haidt, 1999). As with trustworthy judgments of neutral faces, affective judgments (e.g., like/dislike) about faces also occur as quickly as 160ms after presentation (e.g., Pizzagalli et al., 2002). Moreover, emotional facial displays convey information that can influence behavior. For example, facial displays of happiness, such as smiling and laughter, have been shown to promote affiliative tendencies in observers (e.g., Keltner & Bonanno, 1997) and are associated with approach related behavior (e.g., Knutson, 1996). Displays of happiness have also been shown to promote decisions to trust during social interactions. For example, social partners displaying a smile were cooperated with and trusted more than non-smiling social partners during a single interaction version of the Trust Game (Scharelmann et al., 2001). Facial displays of anger, on the other hand, signal to observers to keep their distance (Marsh et al., 2005). People who exhibit angry displays are rated as being less trustworthy (Dunn & Schweitzer, 2005), and are perceived as being less likely to engage in affiliative behavior (Montepare & Dobish, 2003).

Emotions can signal information about a social partner's intentions, which may be useful when deciding whether to trust a social partner during initial encounters. However, a social partner's subsequent reciprocated trustworthy behavior will further shape these impressions and inform future decisions about whether to continue to trust. Indeed, most social encounters involve an ongoing series of exchanges that provide new information that in turn shapes impressions and decisions (Keltner & Kring, 1998). When deciding whether to trust a person, research suggests that we continue to modify our decision-making over the course of repeated interactions, tracking changes in a social partner's

behavior. For example, King-Casas and colleagues (2005) had participants play an repeated version of the Trust Game with another person, with the amounts invested and returned by each player presented in real time. The amount invested for each trial was dependent on the previous trial's return, with larger returns leading to greater investment on the next trial.

Although changes in real-time trustworthy behavior constitute an important type of information used in decision-making, other characteristics of a social partner's behavior are also influential. For example, knowing that a social partner has a history of immoral behavior, regardless of how trustworthy their current behavior may be, was associated with less participant trust during an repeated version of the Trust Game (Delgado, Frank, & Phelps, 2005). These studies highlight the conundrum presented by our example of the smiling used car salesperson. Is his history of untrustworthy behavior salient enough to look past that smiling display? Or, is the display of happiness enough to override warnings about the salesperson's intentions?

To date, studies have focused on manipulating social partner facial displays *or* trustworthy behavior to understand the relative influence of each factor in decision-making. However, to gain a better understanding of the mechanisms underlying decisions to trust, we need to understand how changes in *both* facial displays and behavior converge to shape social decision-making. By simultaneously investigating changes in facial display and behavior, we can examine what happens when these two factors are congruent (e.g., smiling and trustworthy behavior) or incongruent (e.g., angry and trustworthy behavior), expanding our knowledge of how we reach decisions about whom to trust.

Only one study that we are aware of has investigated facial features and trustworthy behavior simultaneously. Chang et al (2010) asked participants to play an repeated version of the Trust Game with four simulated players represented by pictures of neutral faces that had been previously rated as high or low in trustworthiness. Player behavior was manipulated so that simulated players consistently (80%) or inconsistently (20%) reciprocated participant trust. Results indicated that while participant decisions of trust were initially based on simulated player facial features, subsequent decisions were associated with the player's behavior. That is, greater investments were made in players that reciprocated participant trust, regardless of whether that player's face was high or low in trustworthiness.

Although Chang et al.'s findings suggest that trustworthy behavior may be more influential than trustworthy facial features in deciding whom to trust over repeated interactions, two key questions remain unanswered. First, the amount returned by each simulated player was constant from trial to trial, thus failing to capture the variability more typical in real-life social exchanges. Second, the pictures used to represent the players in this study were emotionally neutral. Given the evidence for the role of emotion in decision-making, it remains unclear whether displays of happiness and anger might override tendencies to make decisions based on a player's prior trustworthy behavior. By including smiling or scowling rather than emotionally neutral faces in this study, we sought to examine how emotion impacts decision making in an repeated version of the

Trust Game. In particular, we investigated how congruency and incongruency between emotional display and behavior influence decisions of trust over the course of repeated interactions. This, coupled with varied return rates by the simulated players, created conditions that, more closely approximate real-life interactions.

The Present Study

In the present study, we sought to investigate how emotion and behavior shape decision-making, such as whether to trust the sales pitch of a smiling, yet unscrupulous car salesman. Participants played one of two repeated versions of the Trust Game during which they interacted with four simulated players. In one version, called the Behavior First (BF) version, participants first learned the behavior of each player through repeated interactions. In the other version, called the Face First (FF) version, participants learned the behavior of each player while simultaneously seeing a picture with an emotional display. This design allowed us to investigate how learning player behavior with or without concurrent emotional displays influenced participant decision-making. In addition, by having the two versions, we were able to investigate the distinct as well as combined contributions of emotional displays and behavior over the course of repeated interactions in making decisions to trust.

We tested several hypotheses. First, we expected that participants would invest more in trustworthy players and less in untrustworthy players, with player trustworthiness indicated by the amount each simulated player returned. Second, we expected that facial displays would influence decisions to trust in initial encounters. Specifically, we predicted that participants assigned to the FF version would invest more in trustworthy players with happy emotional displays compared to the amount invested by participants assigned to the BF version in trustworthy players with no accompanying facial display. Similarly, we expected that participants assigned to the FF version would invest less in untrustworthy players with angry displays compared to the investment of participants assigned to the BF version who interacted with untrustworthy players showing no concurrent displays of anger. In other words, the information signaled by emotional displays would be associated with trust (i.e., investment behavior) above and beyond player behavior during these initial encounters.

Following repeated interactions with simulated players, however, we predicted that when emotional display and player behavior were incongruent (e.g. happiness with untrustworthy behavior), participants' decisions to trust would be more associated with the simulated player's behavior. Specifically, following repeated interactions with players of each Trust Game version, we expected that participants would invest more in trustworthy and less in untrustworthy players, regardless of whether their current emotional display or behavior was incongruent with their established pattern of reciprocation. Finally, we examined whether men and women would differ in their decisions to trust. Recent evidence suggests that women are more risk averse than men (see review by Croson & Gneezy, 2009 and meta-analysis by Byrnes, Miller, & Schafer, 2001), particularly in specific domains, such as ethics and finances (Figner & Weber, 2011), suggesting that women may be less likely than men to trust others in this type of

paradigm. However, studies using behavioral economic paradigms to investigate social decision-making have typically not found, or at least not reported, gender differences in risk preference.

Methods

Participants

Seventy-two undergraduate students (36 men, 36 women) from the University of California, Berkeley were randomly assigned to one of two versions of the repeated Trust Game. The mean participant age was 20.51 years (SD 2.98). The sample was predominantly Asian (63.9%), followed by Caucasian (23.6%) and Hispanic (6.9%). Students received partial course credit in a psychology class upon completing the study procedures.

Procedure

After providing informed consent, participants were told that they would be playing a computer game with other people. Participants were led to believe that the other players were real people when in fact each player was simulated, behaving according to a pre-determined pattern. Participants were told they would see pictures of the other players during some of the game trials. To increase believability, each participant had his or her picture taken, ostensibly to be incorporated into the game for the other players to see.

Participants played a repeated version of the Trust Game, created using E-Prime 2.0 software and presented on a Dell desktop computer with an 18" monitor. On each trial, participants had the option of investing from 0 to 10 points in a player. The invested amount was then quadrupled, and the other "players" would then return an amount of the quadrupled sum to the participant (Figure 1 shows an example trial). Participants played against four simulated players (labeled Player 1, 2, 3, or 4). Player behavior, similar to a previous study (van t' Wout & Sanfey, 2008), was predetermined so that two players were trustworthy (i.e., their average return was double the initial amount invested), and the other two players were untrustworthy (i.e., their average return was half of the initial amount invested). For example, a trustworthy player given 6 points by a participant would return an average of 12 of the 24 possible points (the 6 points were quadrupled to 24). By contrast, an untrustworthy player would return an average of 3 of the 24 points. The average amount of points participants invested on a trial, ranging from 0 to 10, represented the degree of trust placed in the simulated players. The total amount of points a participant received did not accumulate across trials but rather was reset after each player interaction.

To further promote optimal performance in the Trust Game, participants were told that they would earn a monetary incentive equal to the sum of six randomly selected trial outcomes. To compute the incentive total, each point above the original ten that the participant was initially given would result in $.10. Thus, a player return of 23 points would yield a payoff of $1.30 for that trial (13 points above the original 10 multiplied by $.10). However, in order to make the incentives equivalent for all participants, everyone received $6 regardless of actual trial outcome. After the Trust Game procedures were

completed, participants were asked two questions (realness of the interactions, believability of the cover story) to assess whether they believed the cover story using a 1 (not at all) to 5 (very much so) Likert scale. A full debriefing interview explained the rationale for deception in detail.

Trust Game Versions

Participants were randomly assigned to one of two versions of the Trust Game, called Behavior First (BF) or Face First (FF). Each version had two blocks; four different simulated players appeared 16 times in the first block (64 trials) and 16 times in the second block (64 trials) for a total of 128 trials. Two key variables distinguished the two game versions. The first variable was the *presence of a facial display* on simulated players in either the second block (BF version) or first block (FF version). The second variable was the *congruence and incongruence of facial display and player behavior*, described in more detail below. Trials with facial displays included simulated players with either a happy or angry facial expression. Table 1 presents details for each version.

Facial Stimuli

For both versions of the game, we used four pictures (2 men, 2 women) displaying either happiness or anger from the NimStim Facial Stimulus Set (Tottenham et al., 2008). Each picture was rated by an independent sample (n=99) to ensure equivalence in intensity, attractiveness, and trustworthiness within valence. That is, pictures were selected such that the two angry pictures were similar in these domains and the two happy pictures were similar in these domains. The mean ratings were as follows, with each rating being made using a 1 (not at all) to 7 (very much so) Likert scale: trustworthiness (angry: 2.01/happy: 5.04), attractiveness (angry: 1.93/happy: 4.15), and intensity (angry: 5.5/happy: 3.2).

Behavior First Version

During the first block (64 trials) of the Behavior First (BF) version, participants played with four simulated players labeled by player number only (1, 2, 3, or 4) with no concurrent facial display (see Table 1). Players 1 and 3 exhibited trustworthy behavior (i.e., returned on average double the invested amount from participants); players 2 and 4 exhibited untrustworthy behavior (i.e., returned on average less than half the invested amount from a participant) At the end of the first block, participants rated the four simulated players' trustworthiness and likeability using a 1 (not at all) to 7 (very much so) Likert scale (see Table 2). In the second block (64 trials) of the BF version, participants played with the same four players identified by the same number as in the first block. However, participants now also saw a picture of each player expressing either a happy or angry facial expression. With the addition of facial displays during the second block of the BF version, we created two conditions: behavior/display congruence (Player 1: trustworthy/happy player, Player 4: untrustworthy/angry player) and behavior/display incongruence (Player 2: untrustworthy/happy player, Player 3: trustworthy/angry player). At the end of the second block, participants again rated player trustworthiness and

likeability.

Face First Version

During the first block (64 trials) of the Face First (FF) version, participants played with four simulated players depicted by an emotional facial display and number (also in Table 1). The pairing of emotional display and player behavior was congruent such that two trustworthy players (Players 1, 3) were paired with happy faces and two untrustworthy players (Players 2, 4) were paired with angry faces. Each player was paired with the same picture throughout the first block. At the end of the first block, participants rated player trustworthiness and likeability (see Table 2). In the second block of the FF version (64 trials), participants played with the same four simulated players, but they no longer saw the pictures that were displayed in the first block. Instead, each player was represented by number only (1, 2, 3, or 4). In addition, , player behavior changed in the second block such that two players' behavior was congruent with behavior in the first block (Players 1, 4) and two players' behavior was incongruent with behavior in the first block (Players 2, 3). Specifically, one player who was trustworthy and represented by a happy display during the first block became untrustworthy (Player 3) and another player who was previously untrustworthy and paired with an angry face became trustworthy (Player 2; see Table 1). This manipulation allowed us to investigate incongruence created by behavior change, rather than the addition of an emotional display, influenced decisions of trust.

During the first block of the FF version and the second block of the BF version, players displayed the same facial emotion throughout the block. However, facial displays were randomly assigned to player behavior across participants such that the same face was paired with different predetermined behavior (i.e., trustworthy, untrustworthy) across participants. Further, to investigate potential effects of player gender on participant decision-making, the pairing of player number and picture were pseudo-randomized so that one gender would be assigned the congruent condition and the other the incongruent condition. Finally, to control for any effect of order, the sequence in which players appeared was pseudo-randomized across participants so that the same player never appeared twice in a row.

Data Analysis Plan

We first examined whether participants assigned to the different Trust Game versions differed on any demographic characteristics. In addition, we examined whether gender of simulated player and order interacted with any other study variables. We tested our main study hypotheses using multiple measures, including participant investment *behavior* and participant self-reported *experience* of their interactions with simulated players. Gender was included in all analyses to investigate whether trust (i.e., investment behavior) was different for men and women. To test our first hypothesis, we conducted a paired-samples t-test to examine the difference between the average amount invested in trustworthy and untrustworthy players during the first block of the BF version. We chose this approach since this block did not include emotional displays, and thus differences in

participant investment would reflect an understanding of player behavioral patterns. We also tested this hypothesis by examining whether participants' ratings of player trustworthiness and likeability for trustworthy and untrustworthy players were consistent with participants' investment behavior.

To test our second hypothesis about greater trust being given to smiling trustworthy players compared to trustworthy players with no smile, and less trust being given to angry untrustworthy players compared to untrustworthy players with no scowl, we conducted a 4 (Player: 1, 2, 3, 4) x 2 (Version: BF, FF) x 2 (Gender: Men, Women) MANOVA for the first block only. By comparing the first blocks of the two versions, we were able to examine the relative contributions of each emotional display (FF version, Block 1) above and beyond player behavior (BF version, Block 1). Similarly, we examined whether participants in the BF versus FF versions differed in their ratings of player likeability and trustworthiness following the first block of interactions.

To test our third hypothesis, we conducted separate 4 (Player: 1, 2, 3, 4) x 2 (Block: First, Second) x 2 (Gender: Men, Women) MANOVA's for each Trust Game version. Analyzing each version independently allowed us to investigate whether the addition of congruent or incongruent emotional displays (BF version) or the change in behavior (FF version) differentially impacted participant's decisions to trust. We also examined whether participant's ratings of trustworthiness and likeability would be associated with current player behavior. For all MANOVAs, Greenhouse-Geisser corrections were used when assumptions for sphericity were not met, and corrected *p* values are reported. Corrected *t*-values and two-tailed *p*-values are reported when Levine's test for equality of variance was not met. Effect sizes are reported as partial eta squared (η_p^2).

Results

There were no differences between the participants assigned to the BF or FF versions in education (t (70) = .36; p = .72) or age (t (70) = .67; p = .51). Additionally, participant race (F (6, 63) = .52; p = .80), the order of player interactions (F (8, 61) = .52; p = .84), and player gender (F (1, 68) = 1.62; p = .19) had no effect on investment behavior, thus they were excluded from further analyses.

Previous studies that have used facial stimuli to represent game players often assume (or at least do not report) that participants believed the cover story. Ratings of how real the interaction felt (average of 2.47 on a 5-point scale) and how believable the cover story about the simulated players was (average of 2.92 on a 5-point scale) suggest that participants did not fully believe nor disbelieve that the interactions were real and with actual players. Of importance, these ratings did not differ between study versions and were not related to task performance.

Our first hypothesis, that participants would invest more in trustworthy compared to untrustworthy players during the first block of the BF version, was supported (t (35) = 9.57; p < .00). That is, participants invested more in trustworthy (M =6.94, SD=1.9) compared to untrustworthy players (M=4.58, SD =1.7; t (35) = 9.57, p < .00), indicating that participants varied their investment behavior depending upon whom they were interacting with. This finding was corroborated by participant ratings of trustworthiness

and likeability of the simulated players at the end of the first block (see Table 2). That is, participants rated trustworthy players (i.e., players 1 and 3) as being more trustworthy (t (35) = 8.33; $p < .00$) and likeable (t (35) = 6.40; $p < .00$) than the untrustworthy players (i.e., players 2 and 4).

Next, we tested whether smiling trustworthy players (players 1 and 3 in the FF version) were more trusted than trustworthy players with no concurrent facial display (players 1 and 3 in BF version), and whether scowling untrustworthy players (players 2 and 4 in FF version) were trusted less than untrustworthy players with no concurrent facial display (players 2 and 4 in BF version) during initial encounters. Examining only the first block of both Trust Game versions, a 4 (player) x 2 (version) x 2 (gender) repeated measures MANOVA revealed a main effect for player (F (3,66) = 107.79; $p <$ .00; $\eta_p^2 = .83$) and a main effect for gender (F (1,68) = 3.97; $p = .05$; $\eta_p^2 = .06$). Overall, participants trusted (i.e., invested more in) trustworthy players more than untrustworthy players, and women trusted less (i.e., invested less) than men. Key to our hypothesis, the Player x Version interaction was also significant (F (3,66) = 6.83; $p < .00$; $\eta_p^2 = .24$). As shown in Figure 2, independent sample t-tests indicated that participants in the FF version invested significantly less in untrustworthy players than participants in the BF version *(t* (70) = 3.43; $p < .00$), indicating that player displays of anger alongside untrustworthy behavior were associated with lower investment amounts than untrustworthy behavior alone. However, participants in the FF version did not invest significantly more in trustworthy players than those in the BF version (t (70) = -.60; $p = .55$) suggesting that the addition of happy displays was not associated with greater trust. Thus, we found only partial support for this hypothesis. Further, these observed differences in investment behavior between the BF and FF versions were not reflected in the trustworthiness and likeability ratings made at the end of the first block in each version. That is, trustworthy and untrustworthy players were rated similarly, regardless of whether they were accompanied by an emotional display.

The Player X Gender interaction was also significant (F (3,66) = 3.38; $p = .02$; η_p^2 = .13). Follow-up t-tests indicated that women assigned to both Trust Game versions invested less in trustworthy players compared to men, (t (70) = 3.14; $p < .00$), but did not differ from men in their investment in untrustworthy players *(t* (70) = .42; $p = .65$). No other main effects or interactions were significant.

Since the creation of incongruent conditions was different in the BF and FF versions, we analyzed each version independently to test our third hypothesis. In the BF version, a 4 (player) x 2 (block) x 2 (gender) repeated measures ANOVA revealed a main effect of player (F (3,32) = 82.4; p < .00; $\eta_p^2 = .89$), indicating that participants invested more overall in trustworthy players than untrustworthy players and a main effect of gender (F (1,34) = 7.29; p = .01; $\eta_p^2 = .18$), indicating that women invested less than men. In addition, the Player x Block interaction was significant (F (3, 32) = 26.78; p < .00; $\eta_p^2 = .72$) and is depicted in Figure 3. Follow-up t-tests indicated that participants invested more in trustworthy players (players1 and 3; t (35) = -2.33; $p = .03$) and less in untrustworthy players (2 and 4; t (35) = 2.81; $p < .01$) during the second block, regardless of whether they were displaying a happy or angry face. In addition, there were no

differences in investment behavior in the second block between the incongruent and congruent conditions for trustworthy (players 1 vs 3: t (35) = 1.23; p = .23) and untrustworthy (players 2 vs 4: t (35) = .90; p = .37). That is, participants did not invest less in a scowling trustworthy player compared to a smiling trustworthy player nor did they invest more in a smiling untrustworthy player compared to a scowling untrustworthy player. Thus, participant' decisions to trust appeared to be influenced more by player behavior than player facial display. These findings were corroborated by ratings completed after the second block of the BF version. Specifically, participants rated trustworthy players as being more trustworthy and likeable regardless of whether their facial display was congruent with their behavior (player 1; trustworthiness: (t (35) = 12.62; p < .00, likeability: t (35) = 10.86; p < .00) or incongruent (player 3; trustworthiness: (t (35) = 9.11; p < .00, likeability: (t (35) = 3.86; p < .00) compared to their untrustworthy counterparts (players 4 and 2, respectively). Interestingly, participants rated the happy untrustworthy player (player 2) as more likeable (t (35) = 2.91; p < .01) than the angry untrustworthy player (player 4) even though investment behavior between towards these two players did not differ, suggesting that smiles may influence perceptions of liking, but not trust.

Next, we investigated whether the removal of concurrent facial displays during the second block of the FF version influenced decision-making. We conducted a 4 (player) x 2 (block) x 2 (gender) repeated measures MANOVA. The player main effect was significant (F (3,32) = 76.87; p < .00; η_p^2 = .88), indicating that participants invested more overall in trustworthy players (1 and 3 in block 1; 1 and 2 in block 2) *(*t (35) = 16.01; p = < .00) than untrustworthy players (2 and 4 in block 1; 3 and 4 in block 2). As predicted, the Player x Block interaction (F (3,32) = 64.21; p < .00; = η_p^2 = .86) was also significant, reflecting the change in players 2 and 3's behavior in the second block compared to the first block. No other main effects or interactions were significant.

As shown in Figure 4, participants invested more in player 2 (t (35) = 10.99; p = < .00) and less in player 3 (t (35) = -10.54; p = < .00) during the second compared to the first block. In other words, participant's seemed to base their decisions to trust on whether a player's current, not past behavior was trustworthy. These findings were also corroborated by ratings of player trustworthiness and likeability. The player with consistent trustworthy across both blocks (player 1) was rated as more trustworthy and likeable than the player who was untrustworthy across blocks (player 4) (trustworthy t (35) = 10.86; p < .00, likeable t (35) = 7.00; p < .00). Further, the player (player 2) whose behavior changed from untrustworthy to trustworthy between blocks was rated as more trustworthy and likeable than the player (player 3) who became untrustworthy between blocks (trust t (35) = 11.20; p < .00, like t (35) = 6.90, p < .00). There were no differences in ratings of trustworthiness or likeability for trustworthy (players 1 and 2) or untrustworthy (players 3 and 4) after the second block. Thus, participant's ratings after the second block reflected the changes in player behavior, not the facial displays from the first block.

Discussion

Facial Emotion and Decision-making

The present study sought to better understand how people use and integrate the information signaled by facial displays with behavior to inform decision-making during social interactions. Our inclusion of emotional displays allowed us to examine how the information signaled by the display was associated with decision-making over the course of repeated interactions. Further, the design of this study allowed us to investigate how congruency or incongruency between the information signaled by a social partner's emotional display and behavior impacted decisions to trust in an attempt to assess what often happens in real life, as in the enigma posed by the used car salesman.

The first question we asked was whether participants would be able to identify trustworthy and untrustworthy players based on their behavior alone. Our findings suggest that participants were able to learn on player behavior, as evidenced by their greater investment in trustworthy compared to untrustworthy players. Similarly, participants rated trustworthy players as more trustworthy and likeable compared to untrustworthy players. Thus, we demonstrated across two modalities (investment behavior, self-report), that participants identified differences in player behavior and used this to inform their subsequent decision-making over the course of repeated interactions.

Our second question addressed whether seeing a player's facial display would influence decisions above and beyond a player's reciprocated behavior. We predicted that happy displays would be associated with greater investment and angry displays associated with lower investment. We found partial support for this hypothesis in that participants invested less in angry untrustworthy players compared to untrustworthy players with no facial display. This finding is consistent with previous studies that have shown angry faces to be rated as less trustworthy (e.g., Dunn & Schweitzer, 2005). However, participants did not rate untrustworthy players with a concomitant angry display as less trustworthy or likable compared to untrustworthy players with no facial display. Thus, while displays of anger were associated in differential investment behavior, they were not associated in differences in participants' experience of the simulated players.

Contrary to expectations, participants did not invest more in trustworthy players exhibiting a smile compared to trustworthy players with no concomitant facial display. Similarly, participants rated trustworthy players as equally likable and trustworthy, regardless of whether the player exhibited a smile. These findings are inconsistent with a previous study that found smiling displays to be associated with greater trust (Scharelmann et al., 2001). However, that study employed a single interaction Trust Game, with participants seeing each display only one time compared to the sixteen times in our study. Thus, our findings suggest that over the course of repeated interactions, displays of anger, but not happiness, continue to provide information that influences decisions to trust.

Another possible explanation for the effect of angry compared to happy facial displays during initial encounters could be the increased salience and evoked arousal of anger. Brain imaging studies of negatively valenced emotional stimuli, such as angry faces, strongly activate the amygdala, a brain area implicated in attentional allocation (Phelps, 2006), perceptual saliency (Anderson et al., 2001) and an integral part in how

affective information influences decision-making (Winkielman et al., 2007; Ernst & Paulus, 2005). While positively valenced emotions, such as happiness, also activate this region (Hamann et al., 2002), studies have shown greater habituation in this area of the brain to "safety" versus "threat" based emotions (Wright et al., 2000). Although not a focus of this study, our findings showed that participants continued to invest less, on average, across the entire first block when presented with an angry player, but did not show a similar pattern for smiling displays. Decreased evoked arousal over the course of repeated exposures to happy compared to angry displays may have contributed to participants not placing greater trust in smiling trustworthy players.

Our third question addressed scenarios where a person's behavior is incongruent with their facial display, as in the example of the used car salesman. We expected that participant's decisions to trust would be more associated with a player's reciprocated behavior, even when that behavior and facial display were incongruent (i.e. happy face but untrustworthy behavior). This is indeed what we found, as participants invested more in a trustworthy player paired with an angry face and less in an untrustworthy player paired with a happy face. Ratings of trustworthiness and likeability mirrored participants investment behavior in that participants rated trustworthy players, regardless of whether they displayed happiness or anger, as more likable and trustworthy compared to untrustworthy players.

We also investigated what happens when incongruency was created by changes in player behavior, such as when a formerly trustworthy player became untrustworthy. Again, we expected a player's current rather than their former behavior would influence that decision-making. Consistent with this notion, participants invested more in currently trustworthy and less in currently untrustworthy players, suggesting a reliance on current behavior for decisions of trust. This was corroborated by ratings of trustworthiness and likeability, which were higher for players with trustworthy players. Support for this hypothesis suggests two things. First, if a social partner's behavior changes, decision-making also updates to reflect the current behavior. Second, if a social partner's behavior remains consistent, but their facial display changes, decision-making will discount the display in favor of behavior. Taken together, these findings point to the dynamic nature of trust. That is, trust develops by incorporating information acquired over the course of repeated interactions to influence current decisions. We extended prior research (Chang et al., 2010; King-Casas et al., 2005) by showing that even when emotional displays are used instead of neutral, current player behavior appears to still be the index for guiding decision-making. That is, over the course of repeated interactions, people place their trust in someone who exhibits trustworthy behavior, regardless of how intense or what information is being signaled by their facial display.

Consistent with previous findings on gender differences in risk aversion (Croson & Gneezy, 2008), we found that women invested less overall than men. As mentioned earlier, previous research has shown that women, compared to men, are less prone to take risks in certain domains, including finances (Figner & Weber, 2011). In the context of the Trust Game, a larger investment can lead to greater payoffs, but a larger investment is also a riskier decision due to the possibility of receiving fewer points in return. Thus,

investing more points in a trustworthy player, despite resulting in a positive return on most trials, can be construed as a riskier decision. Our data showing that women invested less overall, specifically for trustworthy and not untrustworthy players, is consistent with previous literature regarding gender differences in risk taking.

This study had some limitations. First, we included only two emotional displays. We chose these emotions because of the evidence indicating their respective influences on perceptions of trust (Dunn & Schweitzer, 2005; Scharelmann et al., 2001) and because these particular emotional displays have been shown to elicit approach or avoidant behavior in an observer (Knutson, 1996). However, other emotions, even emotions of similar valence, can signal information that may differentially influences decision-making during social interactions (Lerner & Keltner, 2000). For example, the degree to which a display is socially engaging (Kitayama, Mesquita, & Karasawa, 2006) or self-conscious (Tracy & Robins, 2004) may provide alternative conceptual frameworks for investigating how emotion influences social decision-making.

An additional interesting direction to take in future research would be to include low arousal emotional displays. Indeed both happiness and anger are high arousal emotions (Russell, 1980), and it remains unclear whether lower arousal displays might be differentially associated with decisions to trust. Another limitation of our study design is the lack of a neutral emotion condition. Although neutral faces can be perceived as being negatively valenced (Lee et al., 2008), having a neutral condition would have provided an additional comparison for the emotional faces and served as a baseline for determining the saliency of the information signaled by the valenced facial emotions.

The facial stimuli that we used in this study, while balanced within emotion on various attributes, were all Caucasian. Although there was no effect of participant race or interaction between player and participant race, using only Caucasian stimuli limited our ability to investigate the role of race or ethnicity in decision-making. Future studies should include a more diverse sample of facial stimuli to further our understanding of how culture shapes decision-making during social interactions (Weber & Morris, 2010). Finally, despite our efforts to create a design more akin to actual social situations, computer interactions lack the complexity and unpredictability found in real world social exchanges. However, our findings demonstrate that tasks like the Trust Game can be modified to more closely emulate an in-person social interaction, while maintaining the experimental control afforded by such paradigms. One interesting direction would be the inclusion of dynamic versus static stimuli, which may increase the embodiment of the emotional information being signaled in the observer (Sato & Yashikawa, 2006) and increase the social nature of the exchange. Continued investigation of how emotion interacts with social factors and contexts will expand and deepen our understanding of social interactions (Fischer & van Kleef, 2010).

Our findings have implications for everyday decision-making during social interactions. In the case of the used car salesman, our data suggest that people are more likely to trust the reviews describing his past shenanigans than his smiling face, and thus leave the lot, and seek a vehicle elsewhere. Further, as the venue for many social exchanges shifts from in-person to online, the quality and quantity of information

available to inform our decision-making changes as well, with first impressions being made in the absence of actual behavior. For example, reading a potential dating partner's online profile might reveal involvement in ongoing volunteer work alongside pictures of a smiling, seemingly trustworthy individual. Initial positive impressions will be associated with greater trusting and decisions to pursue future interactions. After repeated interactions, however, our data suggest that decisions to trust would be updated based on behavior that is revealed across repeated interactions, no matter how many smiling photos may be posted.

In summary, the present study sought to investigate how facial emotion and behavior influence decision-making during repeated social interactions. Our findings add to the existing emotion and decision-making literature in two main ways. First, facial displays of anger can influence decisions of trust during initial interactions with a person. Second, with repeated interactions, decision-making is continually updated to reflect changes in a social partner's pattern of behavior, even if this pattern is in conflict with the information being signaled by their facial display. These results improve our understanding of how facial displays and behavior influence decisions in the context of repeated social interactions.

References

Anderson, A.K. & Phelps, E.A. (2001) Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411*, 305-309.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*, 122-142.

Byrnes, J.P., Miller, D.C., & Schafer, W.D. (1999) Gender differences in risk-taking: A meta-analysis. *Psychological Bulletin, 125*, 367-383.

Camerer, C. (2003). *Behavioral Game Theory.* New York: Russell Sage Foundation

Chang, L.J., Doll, B.B., van t' Wout, M., Frank, M.J., Sanfey, A.G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology, 61*, 87-105

Croson, R. & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature, 47*, 448-474.

Delgado, M.R, Frank, R.H, & Phelps, E.A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature Neuroscience, 8*, 1611-1618.

Dunn, J.R. & Schweitzer, M.E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology, 88,* 736-748.

Ernst, M. & Paulus, M.P. (2005) Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. *Biological Psychiatry, 58*, 597-604.

Figner, B. & Weber, E.U. (2011). Who takes risks and why?: Determinants of risk taking. *Current Directions in Psychological Science, 20*, 211-216.

Fischer, A.H., & van Kleef, G.A. (2010) Where have all the people gone? A plea for including social interaction in emotion research. *Emotion Review, 2*, 208-211.

Hamman S.B., Ely, T.D., Hoffman, J.M., & Kilts, C.D. (2002). Ecstasy and agony: Activation of the human amygdala in positive and negative emotion. *Psychological Science, 13*, 135-141.

Keltner, D., & Bonanno, G.A. (1997). A study of laughter and dissociation: Distinct correlates of laughter and smiling during bereavement. *Journal of Personality and Social Psychology, 73*, 687-702.

Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion, 13,* 505-521.

Keltner, D., & Kring, A.M. (1998). Emotion, social function, and psychopathology. *Review of General Psychology, 2*, 320-342.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C., Quartz, S., & Montague, P.R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*, 78-83.

Kitayama, S., Mesquita, B., & Karasawa, M. (2006). Cultural affordances and emotional experience: Socially engaging and disengaging emotions in Japan and the United States. *Journal of Personality and Social Psychology, 91*, 890-903.

Knutson, B. (1996). Facial expressions of emotion influence interpersonal trait inferences. *Journal of Nonverbal Behavior, 20*, 165-182.

Lee, E., Kang, J.I., Park, I.H., Kim, J.J., & An, S.K. (2008) Is a neutral face really evaluated as being emotionally neutral. *Psychiatry Research, 157*, 77-85.

Lerner, J.S. & Keltner, D. (2000) Beyond valence: Toward a model of emotion-specific influences on judgment and choice. *Cognition and Emotion, 14*, 473-493.

Marsh, A.A., Ambady, N., & Kleck, R.E. (2005). The effects of fear and anger facial expressions on approach and avoidance related behaviors. *Emotion, 5*, 119-124.

Montepare, J.M. & Dobish, H. (2003). The contributions of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal Behavior, 27*, 237-254.

Phelps, E.A. (2006) Emotion and cognition: Insights from studies of the human amygdala. *Annual Reviews of Psychology, 57*, 27-53.

Pizzagalli, D.A., Lehmann, D., Hendrick, A.M., Regard, M., Pascual-Marqui, R.D., & Davidson, R.J. (2002). Affective judgments of faces modulate early activity (~ 160 ms) within the fusiform gyri. *Neuroimage, 16,* 663-677.

Rilling, J.K., & Sanfey, A.G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology, 62*, 23-48.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161-1178.

Sato, W. & Yoshikawa, S. (2006) Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition, 104*, 1-18.

Scharlemann, J.P.W., Eckel, C.C., Kacelnik, A., & Wilson, R.K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology, 22*, 617-640.

Todorov, A., Pakrashi, M., & Oosterhof, N. (2009). Evaluating faces of trustworthiness after minimal time exposure. *Social Cognition, 27*, 813-833.

Tottenham, N., Tanaka, J., Leon, A.C., McCarry, T., Nurse, M., Hare, T.A., Marcus, D.J., Westerlund, A., Casey, B.J., Nelson, C.A. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, 168(3):242-9.

Tracy, J.L. & Robins, R.W. (2004) Putting the self into self-conscious emotions: A Theoretical model. *Psychological Inquiry, 15*, 103-125.

van t' Wout, M. & Sanfey, A.G. (2008) Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*, 796-803.

Weber, E.U. & Morris, M.W. (2010) Culture and judgment and decision making: The constructivist turn. *Perspectives on Psychological Science, 5*, 410-419.

Winkielman, P., Knutson, B., Paulus, M., & Trujillo, J.L. (2007) Affective influence on judgments and decisions: Moving towards core mechanisms. *Review of General Psychology, 11*, 179-192.

Wright, C.I., Fischer, H., Whalen, P.J., McInerney, S.C., Shin, L.M., & Rauch, S.L. (2000) Differential prefrontal cortex and amygdala habituation to repeatedly presented emotional stimuli. *NeuroReport, 12*, 379-383.

Figure Captions.

Figure 1. The first screen showed the player number. During some blocks (as shown in this example), players were also represented by an emotional display. The second screen prompted participants to make their investment. The third screen displayed the quadrupled point total. The fourth screen showed the amount returned by the player as well as both participant and player totals for that trial.

Figure 2. This figure shows the average amount of points (0-10) invested in each player during the first block of the BF and FF versions. Participants invested significantly less points in untrustworthy player accompanied with angry emotional displays (i.e., Players 2 and 4 in the FF version) than untrustworthy players with no concurrent angry display.

Figure 3. This figure shows the average amount of points (0-10) invested in each player during the first and second blocks of the BF version. After learning player behavior behavior in the first block, facial displays that were congruent or incongruent with player behavior were added during the second block. Participant investment in the second block reflected player behavior with significantly more points given to trustworthy players (1 and 3) and fewer points to untrustworthy players (2 and 4) compared to the first block.

Figure 4. This figure shows the average amount of points (0-10) invested in each player during the first and second blocks of the FF version. During the first block, participant's learned player behavior with concurrent facial displays. In the second block, displays were removed and the behavior of two players (2 and 3) changed. Participants were able to pick up on the changes in the behavior of players 2 and 3, but investment was in trustworthy and untrustworthy players were not different across blocks.

Table 1

Overview of Study Design with Player Behavior and Emotional Displays for each Block and Version

	Behavior First		Face First	
	Block 1	Block 2	Block 1	Block 2
Player 1	Trustworthy	 & Trustworthy	 & Trustworthy	Trustworthy
Player 2	Untrustworthy	 & Untrustworthy*	 & Untrustworthy	Trustworthy*
Player 3	Trustworthy	 & Trustworthy*	 & Trustworthy	Untrustworthy*
Player 4	Untrustworthy	 & Untrustworthy	 & Untrustworthy	Untrustworthy

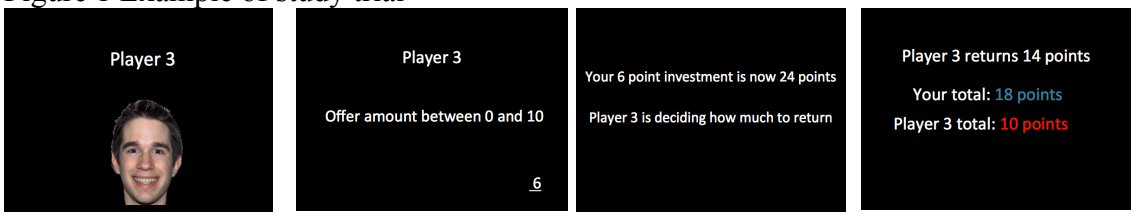
Note. * = Incongruent conditions
Note. Faces used in this table are for the purpose of demonstration and were not the actual facial stimuli used during the study.

Table 2
Trustworthiness and likeability ratings in each version by block

	Behavior First		**Face First**	
	Block 1	Block 2	Block 1	Block 2
	M (SD.)	M (SD)	M (SD)	M (SD)
Player 1	Trust = 4.91 (1.46)	Trust = 5.74 (1.10)	Trust = 4.37 (2.03)	Trust = 5.63 (1.17)
	Like = 5.11 (1.51)	Like = 5.80 (1.02)	Like = 4.71 (1.96)	Like = 5.23 (1.60)
Player 2	Trust = 2.76 (1.18)	Trust = 2.23 (1.11)	Trust = 3.00 (1.59)	Trust = 5.69 (1.00)
	Like = 3.18 (1.53)	Like = 3.63 (2.00)	Like = 3.34 (1.78)	Like = 5.31 (1.41)
Player 3	Trust = 5.14 (1.35)	Trust = 5.29 (1.45)	Trust = 4.46 (1.72)	Trust = 2.34 (1.40)
	Like = 5.09 (1.53)	Like = 5.11 (1.18)	Like = 4.20 (1.55)	Like = 2.66 (1.78)
Player 4	Trust = 2.89 (1.61)	Trust = 2.00 (1.11)	Trust = 3.26 (1.21)	Trust = 2.17 (1.40)
	Like = 2.97 (1.52)	Like = 2.26 (1.67)	Like = 3.40 (1.54)	Like = 2.51 (1.84)

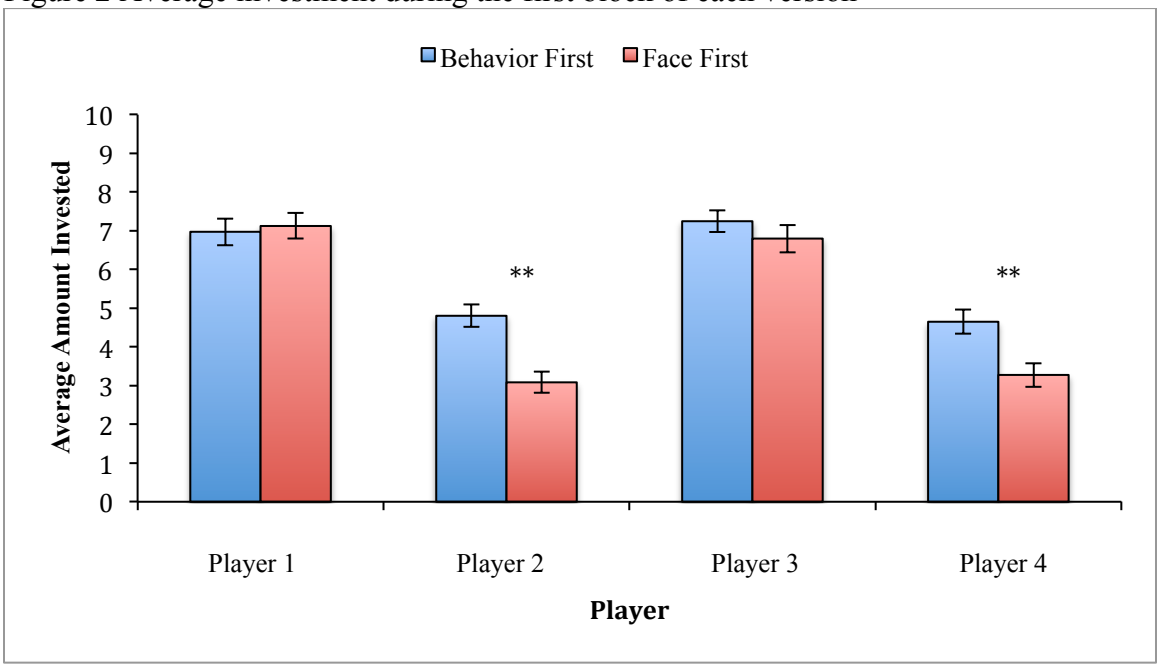
M = mean, SD = standard deviation, Trust = trustworthiness, Like = likeability
Note. All ratings were made using a 1 (not at all) to 7 (very much so) Likert scale

Figure 1 Example of study trial



3000 ms | Untimed | 2000-4000 ms | 4000 ms

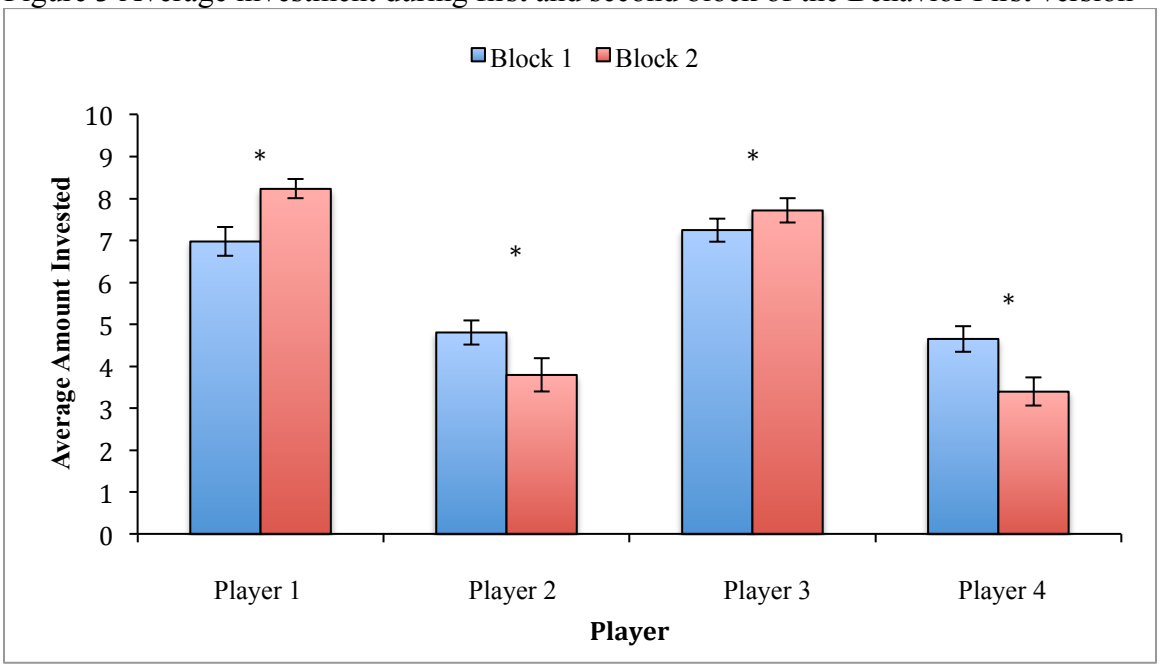
Figure 2 Average investment during the first block of each version



Note: Players 1 and 3 in the BF version were trustworthy; Players 1 and 3 in FF version were smiling and trustworthy; Player 2 and 4 in BF were untrustworthy; Players 2 and 4 in FF version were angry and untrustworthy.

** $p < .01$

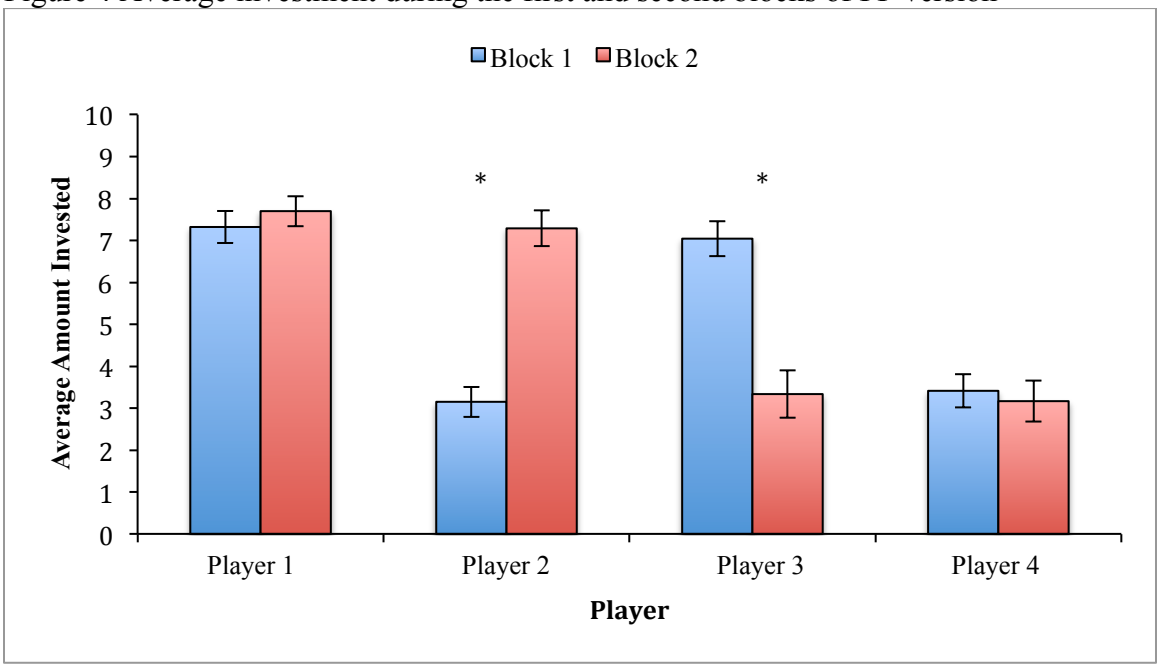
Figure 3 Average investment during first and second block of the Behavior First version



Note: Players 1 and 3 were trustworthy and Players 2 and 4 were untrustworthy in both blocks. Players 1 and 2 displayed smiles in block 2; Players 3 and 4 displayed anger. Thus, players 2 and 3 exhibited incongruence between behavior and facial display in block 2 whereas players 1 and 4 exhibited congruence between behavior and facial display in block 2.

* $p < .05$

Figure 4 Average investment during the first and second blocks of FF version



Note: Players 1 and 3 were trustworthy in block 1; Players 1 and 2 were trustworthy in block 2. Players 2 and 4 were untrustworthy in block 1; Players 3 and 4 were untrustworthy in block 2. Players 1 and 3 displayed smiles in block 1; Players 2 and 4 displayed anger in block 1. Thus, players 2 and 3 exhibited incongruence between behavior and facial display in block 2 whereas players 1 and 4 exhibited congruence between behavior and facial display in block 2.

* $p < .05$