Conflicted Emotions Following Trust-based Interaction

Eric Schniter and Roman Sheremeta and Timothy Shields

17. August 2015

Online at http://mpra.ub.uni-muenchen.de/66154/
MPRA Paper No. 66154, posted 18. August 2015 05:57 UTC
Conflicted Emotions Following Trust-based Interaction

Eric Schniter
Economic Science Institute, Chapman University, One University Drive, Orange, CA 92866, U.S.A.*

Roman M. Sheremeta
Weatherhead School of Management, Case Western Reserve University, 11119 Bellflower Road, Cleveland, OH 44106, U.S.A.

Timothy W. Shields
Argyros School of Business and Economics & Economic Science Institute, Chapman University, One University Drive, Orange, CA 92866, U.S.A.

Abstract:
We observed reports of conflicted (concurrent positive and negative) emotions activated after interactions in the Trust game. Our analyses reveal that activation of 20 emotional states following trust-based interaction is better explained by predictions derived from a multi-dimensional Recalibrational perspective than by predictions derived from two-dimensional Valence and Arousal perspectives. The Recalibrational perspective proposes that emotions are activated according to their functional features – for example, emotions help people achieve short or long-sighted goals by up or down-regulating behavioral propensities, whereas Valence and Arousal perspectives consider simpler hedonic dimensions lacking functional specificity. The Recalibrational perspective is also distinguished from the Valence and Arousal perspectives in that it predicts the possibility of conflicted emotions. We discuss the theoretical implications of having conflicted goals and the economic implications of having conflicted emotions.

Keywords: emotion, affect valence, Recalibrational theory, intrapsychic conflict, Trust game

PsycINFO Classification code: 2360

JEL code: C73, C91, D87

* Corresponding author. Address: Economic Science Institute, Chapman University, One University Drive, Orange, CA 92866, USA. Tel.: (714)628.7272. E-mail address: eschniter@gmail.com
1. Introduction

Both laypeople and theorists tend to view the simultaneous experience of positive and negative emotions (e.g. happiness and sadness) as abnormal. For example, Zautra, Potter, and Reich (1997) demonstrated that laypeople believe happiness and sadness to be opposites and not capable of coexisting. This lay theory has been shown cross-culturally (Almagor & Ben-Porath, 1989; Russell, 1983) and among children (Russell & Bullock, 1985). Likewise, emotion theorists have traditionally viewed positive and negative emotions as two mutually exclusive ends of a continuum and therefore uncommon as co-occurring experiences (e.g. Russell, 1979; Russel & Carroll, 1999).¹ A more recent body of work has challenged this traditional view, suggesting that simultaneous experience of positive and negative emotions is a normal consequence of intrapsychic conflict. The phenomenon of conflicted emotions has been described by various names such as mixed feelings (Kahneman, 1992) emotional ambivalence (Fong, 2006), mixed emotions (Hong & Lee, 2010), and compound emotions (Du, Tao, & Martinez 2014). Behavioral and neurological evidence suggests that the systems involved in positive and negative emotions are functionally and structurally distinct (McClure et al., 2004; Hare et al., 2009) and that these modular systems can be co-activated (Miller 1960).

¹ In circumplex models of emotion, happiness and sadness are depicted as opposite one another, indicating that they should not co-occur (e.g., Watson & Tellegen, 1985; Watson & Tellegen, 1999; Russell & Carroll, 1999; Russell & Feldman-Barrett, 1999). Likewise, in the two dimensional V-shaped “Arousal” models it is assumed that arousal reflects either the intensity of pleasure or displeasure but never both (e.g., Clore, Ortony, & Foss, 1987; Lang, 1994; Kuppens et al., 2012).
The suggestion that conflicted emotions are a regular everyday phenomenon has disturbed lay intuitions and traditions of academic thinking, causing debate (Feldman Barrett & Russell, 1999) and motivating a wave of research to document just how normal and reliably evoked conflicted emotions really are. Research has shown that participants report feeling conflicted emotions in response to bittersweet movies (Larsen & McGraw, 2011), movies with disgusting humor (Hemenover & Schimmack 2007), sentimental periods of transition (Larsen, McGraw, & Cacioppo, 2001), organ donations (Parsi & Katz, 1984), difficult ethical questions (Priester & Petty, 1996), and self-control dilemmas (Ramanathan & Williams, 2007). Hong and Lee (2010) reviewed a number of conflicted emotion experiences from across a wide variety of “consumer contexts” suggesting that they are ubiquitous phenomena with economic relevance. Furthermore, because people appear capable of signaling their subjective conflicted emotional states through facial expression and can recognize these expressions (Rothman, 2011; Du et al., 2014), this topic becomes relevant to studies of interpersonal behavior. Despite the rich literature on conflicted emotions and a broad literature on emotions and cooperation, little attention has been given to the experience of conflicted emotions in the context of cooperation or trust-based interaction.

Motivated to better understand how trust-based interaction can trigger conflicted emotions, we examined reports of 20 emotions by 170 participants after completing a Trust game and learning of its outcome. Using this approach, we investigated whether emotions were experienced in a patterned way that conforms to predictions derived from a Recalibrational perspective (Tooby & Cosmides, 1990; Nesse, 1990; Schniter & Shields, 2013; Schniter & Sheremeta, 2014) or Valence and Arousal perspectives (e.g., Ortony, Clore, & Collins, 1988;
Lang et al., 1993; Cacioppo and Berntson, 1994; Russell & Carroll, 1999). The Recalibrational perspective proposes that specific sets of emotions are activated or aroused in response to problems, and that these sets of emotions are distinguished by their evolved functional design features (positive or negative recalibrations of short- and/or long-sighted programs). For example, Schniter and Sheremeta (2014) have provided evidence that recalibrational emotions triggered by a Trust game’s outcome predict investors’ subsequent trust re-extension and whether trustees will be trustworthy. Evoked in response to problems, these recalibrational emotions are best classified with immediate and integral emotions because they up or down-regulate subsequent behavioral propensities to pursue short or long-sighted goals. Simpler Valence and Arousal perspectives lacking in etiological sophistication predict that emotions should be activated according to a positive-negative affect dimension alone. The Recalibrational perspective additionally considers the short- and long-sighted adaptive goals served by subsets of valenced emotions. As such, the Recalibrational perspective is

---

2 “Programs” is a term borrowed from computational science, referring to neural circuits in the brain/body that process input information and accordingly cause outputs either in the form of regulatory feedback (reused as input by programs) or behavior.

3 Rick & Loewenstein (2008) have described two basic types of emotions: expected emotions and immediate emotions. Expected emotions are those that are anticipated to occur as a result of the outcomes associated with different possible courses of action. Immediate emotions are experienced the moment a decision or event takes place and fall into one of two categories: incidental or integral. The recalibrational emotions that we focus on are emotions experienced immediately as a consequence of trust-based interactions. These immediate recalibrational emotions are not incidental emotions (i.e., triggered by something unrelated to the decision at hand), but rather integral emotions that arise from interaction outcomes (in this case trust-based exchanges) and contribute to an emotional capital that subsequently affects related interaction decisions (subsequent trust-based exchanges that are repeated with the same exchange partners).
distinguished from most Valence perspectives in that it predicts when subsets of valenced emotions should become activated and when emotional conflict occurs.\(^4\)

Our results indicate that, after interactions in the Trust game, participants simultaneously experience activation of positive and negative subsets of emotions consistent with predictions derived from the Recalibrational perspective but not Valence perspectives. We use confirmatory factor analysis to show that emotional responses load onto four sets derived from the Recalibrational model. Indeed, we find that participants report experiencing simultaneous activation of both negative and positive emotions across these sets (e.g. simultaneous guiltiness with contentment or anger with pride). Estimating a structural equation model to compare the fit of models to observed data, we find that the Recalibrational model significantly outperforms the Valence models. In summary, our results support the hypotheses that several distinct sets of emotions are triggered in patterned response to the adaptive problems produced by trust-based interactions and that the experience of conflicted emotions may result.

2. **Theoretical predictions**

2.1 *The adaptive dilemma modeled by the Trust game*

When one is confronted with a dilemma, there is an internal psychic conflict over how to pursue alternative desired outcomes that cannot be simultaneously fulfilled at their maxima.\(^5\)

---

\(^4\) As an extension of the Valence models widely applied to bipolar ratings scales, Cacioppo and Berntson (1994) suggested a bivariate formulation of positively and negatively valenced evaluative processes and attitudes to help explain evidence of the separable activation of positive and negative evaluations seen in behavioral studies. In this study we also evaluate and compare an unrestricted form of the Valence model akin to the bivariate formulation (where interdependence is *unrestricted* between positively and negatively valenced emotions).

\(^5\) For discussion of genetic origins and various manifestations of intrapsychic conflict see Trivers (1997).
We study such a dilemma modeled by Berg, Dickhaut, and McCabe (1995), which we refer to as the Trust game. In our version of the Trust game, an investor first decides how much of a $10 endowment to send (= s) to a paired trustee, with the sent amount tripled, and then the trustee decides how much of the tripled investment, or income, to return (= r) to the investor.

From a short-sighted perspective, the Trust game provides opportunity for gaining available resources. From a long-sighted perspective, the Trust game provides the possibility of developing the foundations for a trust-based exchange relationship that our minds consider a security against income risks associated with luck-based asymmetries (such as resulting from the 50% chance of being the investor in this kind of experiment). We propose that the pursuits of these goals are regulated by evolved short-sighted and long-sighted behavior regulation programs in conflict with one another (Carrillo, 1998; Livnat & Pippenger, 2006; Kurzban, 2010).⁶

In the Trust game, the programs regulating an individual’s propensity to pursue short and long-sighted goals are likely affected by a number of factors.⁷ According to our Recalibrational perspective, the investor’s decision trades off his short-sighted “opportunistic” goal (achieved

---

⁶ Short-sighted programs appear evolved to solve the adaptive problem of competition for limited resources with fleeting availability by encouraging capture of all resources present before they are depleted, foregone, or the possibility of seizing them becomes less certain or riskier. Long-sighted programs appear evolved to solve the adaptive problem of developing reliable trust based exchange relationships: important securities that buffer against resource shortages and times of scarcity associated with risky income. Indeed, laboratory studies have demonstrated that, in response to unsynchronized resource availability among individuals in a common environment, people act pre-disposed to engage in asynchronous trading relationships (Kaplan et al., 2012).

⁷ The calibration of one’s behavior regulating programs will be determined by moods, emotional capital (consequent on past goal accomplishments or forgone achievements), present demands, available outcomes, and belief-dependent emotions based on expectations about a partner (e.g. see Chang et al., 2010, 2011). So, while we expect the output of these programs to show individual differences in degree (i.e., variance in relative strengths of regulatory programs or emotions), we do not expect them to show differences in kind (i.e., direction or dynamics of recalibrational effects) if they exist as reliably developing species-typical adaptations.
Conflicted Emotions Following Trust-based Interaction

with earnings from a kept endowment and a maximally profitable investment) with pursuit of his long-sighted “cooperative” goal (achieved by developing an exchange relationship in which both trust and trustworthiness are maximally demonstrated). Likewise, the trustee, having received a trust-based multiplied transfer of funds from the investor, must decide whether to pursue her short-sighted program’s goal by keeping this income, or else pursue her long-sighted program’s goal. The trustee’s long-sighted program’s goal is to develop a trust-based exchange relationship by returning an amount equal to or greater than what the investor originally sent and thereby demonstrating her trustworthy cooperativeness.

After a Trust game, an individual’s integration of new information (from trust-based decisions and interaction outcomes) triggers the activation of “immediate” positive and negative emotions serving subsequent short- and long-sighted goal pursuits. A novel feature of our model is that it identifies Trust game outcomes predicted to trigger conflicted emotions.

2.2 Description of the Recalibration model

Here we briefly describe a three-stage Recalibrational model of (1) trust-based interaction regulation, (2) emotional activation, and (3) emotional recalibration (of subsequent trust-based interaction regulation), so as to properly contextualize our conceptualization of emotions from a Recalibrational perspective.

In the first stage of a Recalibrational model of trust-based interactions, the relative calibration of “short-sighted”, and “long-sighted” programs is theorized to ultimately determine investor and trustee behavior propensity in a Trust game choice dilemma. The balance of short- and long-sighted program weights determines behavior propensity via a
decision function, where the long-sighted program weakly increases and the short-sighted program weakly decreases amount sent or amount returned, for the investor and trustee, respectively. Thus, the relative power of these programs determines the extent to which an individual’s behavior in a trust-based interaction trades off the short-sighted goal (opportunism) for the long-sighted goal (developing a trust-based exchange relationship).

The second stage of the Recalibrational model describes emotion activation and is the main focus of the study reported here. In this focal part of our model, an individual’s integration of new information (from trust-based decisions and interaction outcomes) triggers the activation of four basic sets of emotions. Emotions are expected to act jointly in these sets based on the recalibrational functions they are believed to share. This second stage of the model also identifies conditions predicted to trigger conflicted emotions. These predictions are generated by the theory that individuals harbor conflicting adaptive goals and that emotions serve these conflicted goals by computationally identifying and responding to the presence of specific adaptive problems emerging from the Trust game. Of particular interest to our study are the “broken trust” problems that can result from unreciprocated investments in the Trust game and the conflicted emotions they trigger. For example, unreciprocated investment can trigger the conflicted emotions pride and anger among investors, and among opportunistic trustees it can trigger conflicted contentment (from having kept all their income) and guilt (for not returning more than the investor’s investment).

8 While the design of our present study can inform us about emotion activation, it cannot provide evidence for how prior calibration of a behavior regulation system affects Trust game decisions, nor can it provide evidence for how triggered emotions actually affect subsequent regulation of trust and trustworthiness.
The third stage of the Recalibrational model describes the directional effects (i.e., “positive” upregulation and/or “negative” downregulation of short-sighted and long-sighted programs) that activated emotions are known to produce. Though these recalibrational effects are untested with this research, Schniter and Sheremeta (2014) provide experimental evidence from Trust games that recalibrational emotions affect trusting and trustworthy behaviors in self and others.

2.3 Proposed recalibrational features of emotions and their predicted activation

Based on our review of the emotion literature and the functional features of recalibrational emotions (Schniter & Shields, 2013; Schniter & Sheremeta, 2014), we consider 20 emotions that cluster into categories or “Sets” (see Table 1) based on constellations of their shared functional features. We chose to classify and predict the twenty emotional states studied because they are frequently used in versions of the one-dimensional Positive and Negative Affect Scale (PANAS) developed by Watson et al. (1988), and predicted by Valence models that we compare to a Recalibrational model. This large battery has also been used in the context of other Trust game studies (e.g. see Kausel & Connolly, 2014), and is comparable among widely used measures of multiple emotional states.

Our functional classification of twenty emotions yields four basic sets for evaluation; each of which contains multiple emotions.9 We expect emotions in a set to be triggered in concert

9 A fifth set containing a single emotion, surprised, is classified as positively or negatively valenced but cannot be adequately tested as a unique factor.
for common functional purposes. We characterize these functions as *positive* and *negative* recalibrations of *short-* and *long*-sighted programs.

Generally, an adaptationist and functional perspective of emotions (e.g., Tooby & Cosmides, 1990; Buck, 1999; Cosmides & Tooby, 2000; Ketelaar, 2006; Schniter & Shields, 2013) argues that emotions facilitate behavioral regulation by recruiting the assistance of a number of psychological, physiological, and behavioral processes that provide either positive or negative feedback used in updating the calibration of conflicting regulatory programs. Positive and pleasant experiences are rewarding and can incentivize continuation of the prior behaviors or interactions that triggered them (Watson et al., 1999; Carver & Scheier, 1990). Negative and unpleasant experiences are costly and motivate a change, whether through behavior reduction, avoidance, or aggression (Gray 1971; Nesse, 1991). Of the set of twenty emotional states, we conjecture that nine [*appreciative, happy, content, cheerful, triumphant, inspired, secure, proud, believable*] are only experienced as pleasant (forming a Positive Affect set), one [*surprise*] could be either pleasant or unpleasant, and ten [*disgusted, jealous, aggravated, frustrated, angry, depressed, sad, embarrassed, ashamed, guilty*] are only unpleasant (forming a Negative Affect set).

When one’s prior actions did not succeed in achieving an adaptive goal negative emotions are triggered to recalibrate one’s regulatory programs (Carver & Scheier, 1990; Baumeister & Heatherton, 1996). Guilt is one such negative emotion that appears triggered exclusively in response to failure of a long-sighted program, such as when one discovers that they have undervalued another’s welfare – potentially harming the relationship (Sznycer et al., 2015). Feeling *guilty* increases the propensity to engage in remedial and cooperative behavior.
Conflicted Emotions Following Trust-based Interaction (Wicker, Payne, & Morgan, 1983). For example, Schniter and Sheremeta (2014) show that trustees who feel guilty or ashamed about their behavior in previous trust-based interactions are likely to produce apologies in an effort to rebuild damaged trust. On the other hand, when one’s prior actions have succeeded in achieving an adaptive goal, we hypothesize that positive emotions are triggered and recalibrate regulatory programs to ensure further achievements. People feel believable and proud – positive emotional states—when they have made decisions that contributed towards cooperative relationships. Schniter and Sheremeta (2014) show that trustees who feel believable for honoring promises and forgoing available opportunism in previous trust-based exchanges are more likely to produce messages reaffirming the success of their recent interaction. These positive social emotions are hypothesized to upregulate the long-sighted program relative to the short-sighted program, so as to further encourage more of the behavior that led to successful cooperation.

Another way that emotions appear designed to function is by affecting others and their subsequent interaction behaviors. For example, consider feeling appreciative. Discovery that another has foregone short-term rewards in the pursuit of a long-term exchange relationship, for example by providing resource or assistance, presents a fortunate relationship building opportunity for the recipient. Appreciation or gratitude can signal one’s favorable valuation of the other and pre-commitment or propensity to cooperate with them (Hirshleifer, 1987; Tooby & Cosmides, 2008), encouraging future trust much in the way that “promises” do (Schniter, Sheremeta, & Sznycer, 2013). Experimental evidence supports this functional account of appreciative and grateful feelings (Tesser, Gatewood, & Driver, 1968; McCullough et al., 2001; Dunn & Schweitzer, 2005; Algoe, Haidt, & Gable, 2008).
Most of the emotions studied appear designed to affect one’s own behavior and the behavior of others. Consider, for example, feeling ashamed and embarrassed following an action. When experienced by an “offender”, shame and embarrassment cause the offender to self-impose negative recalibrations, so as to mitigate the likelihood or costs of reputation-damaging information spreading to others (Sznycer et al., 2012). These self-directed recalibrations should also reduce the likelihood of repeating the shameful or embarrassing actions and, if signaled, may preempt punishment or rejection by angry victims that tend to non-cooperation (Pillutla & Murnighan, 1996). Shame and embarrassment could also mollify a victim’s anger by acting remedially: when rule violators demonstrate that they have subsequently suffered hedonic displeasure yet retain concern for the victims’ welfare, the angry and aggressive responses of offended parties are preempted (Keltner, Young, & Buswell, 1997; de Jong, 1999). While we conjecture that fifteen of the twenty emotional states studied may facilitate the achievement of either short- or long-sighted programs’ goals, we consider five emotional states to exclusively facilitate achievement of the long-sighted program’s goal. Of these we derive two unique sets: a positive Set 2 [proud, believable] and a negative Set 4 [embarrassed, ashamed, guilty]. The positive emotional states that facilitate both short-sighted and long-sighted programs [appreciative, happy, content, cheerful, triumphant, inspired, secure] form the unique Set 1. The negative emotional states that facilitate both short-sighted and long-sighted programs [disgusted, jealous, aggravated, frustrated, angry, depressed, sad] form the unique Set 3.

Our Recalibrational theory of emotions is built around conflicting short-sighted and long-sighted behavior regulation programs that determine an individual’s choices when faced with
Conflicted Emotions Following Trust-based Interaction

decision dilemmas, such as in the Trust game. We propose that recalibrational emotions assess
game outcomes for the purpose of identifying and reacting to successes and failures of the
short-sighted and long-sighted programs in self and other. To evaluate our propositions, we test
whether specific antecedents produced by the Trust game outcomes reliably predict specific
sets of emotional experiences.

Below we provide the set of predictions generated by Valence models and by our
Recalibrational model. Each prediction provides characterization of a relationship between
emotions and Positive Affect (PA) or Negative Affect (NA) sets, as well as a relationship
between PA and NA sets.

2.3.1. Valence model predictions

**P1:** Emotions are positively correlated with PA and NA sets. PA and NA sets are
*independent* with no correlation (= 0) between them.

**P2:** Emotions are positively correlated with PA and NA sets. PA and NA sets are strictly
*interdependent* with negative (= -1) correlation between them. Consistent with a purely
“bipolar” model of valence, reports of simultaneously experienced strong positive emotion
and strong negative emotion are not expected.

**P3:** Emotions are positively correlated with PA and NA sets. PA and NA sets are strictly
*interdependent*, but such interdependence is *unrestricted* between emotions in the PA and
NA sets. While negative correlation is expected between sets, positive correlation between
items in PA and NA sets can also occur.
Conflicted Emotions Following Trust-based Interaction

2.3.2. Recalibrational model predictions

P4: According to the Recalibrational model, Trust game interactions and outcomes (as captured by short- and long-sighted program triggers) trigger the activation of emotions in four basic sets (i.e. Sets 1-4 detailed in Table 1). By extension, we predict the outcomes in a Trust game that are more likely to trigger simultaneous positive and negative emotion activation (i.e., when trust is extended but trustworthiness not).

2.3.3. Recalibrational model versus Valence models

P5: Trust game interactions and outcomes (as captured by short- and long-sighted program triggers) predict the experience of emotions for the 4 testable sets of the Recalibration model better than for the 2 sets (PA, NA) of the Valence model.

We tested each of these predictions using the natural experiment described below.

3. Method

3.1 Participants and sampling procedure

At Chapman University’s Economic Science Institute, we sampled 170 participants (83 males and 87 females) recruited from a campus-wide subject pool consisting primarily of undergraduate students. The number of participants was determined by resource constraint. Participants who had previously participated in trust experiments were not recruited. All participants consented to the procedures of the study, which were approved by Chapman University’s Internal Review Board.
3.2 Natural experiment design

We took a natural experiment approach, conducting a Trust game in which the investor received an endowment of $10 and could send any portion of it to the trustee, with the amount sent tripled (see Appendix A). The trustee then decided how much of the tripled investment, or income, to return (or else keep). We define the following variables observed in our Trust game: endowment \((= e)\), amount sent by investor \((= s)\), amount returned by trustee \((= r)\). Following the Trust game we administered a 20-item emotional status survey (see Appendix B) in which participants reported how much they felt activation of each of 20 emotions (on a five point scale labeled (1) very slightly or not at all, (2) a little, (3) moderately, (4) quite a bit, (5) extremely) as a consequence of their recent game interactions and outcomes.\(^{10}\) The computer software presented each participant the full randomly ordered set of all emotional states listed in Table 1. Using this laboratory implementation of the Trust game that engaged participants in one-shot anonymous economic interactions, followed by a well-established emotional status survey, we investigated whether naturalistically triggered emotional experiences were reported in a patterned and predicted way as a consequence of game outcomes.\(^{11}\)

\(^{10}\) To avoid experimenter demand effects that might result by soliciting reports on only a few select emotional states commonly ascribed to failed trust-based interactions (i.e., anger and guilt) and identified in the literature (e.g., Ketelaar & Au, 2003; Kausel & Connolly, 2014), we constructed a survey of a large array of emotional states, based on the Positive and Negative Affect Scales (PANAS), a self-report measure of positively and negatively valenced affect state activations developed by Watson et al. (1988) that has been demonstrated across large non-clinical samples to be a reliable and valid measure of these states (Crawford & Henry, 2004). Consistent with the moderately high reliability of internal consistency reported previously by Watson et al. (1988) and others (e.g. Jolly et al., 1994; Mehrabian, 1998; Roesch, 1998; Kausel & Connolly, 2014) we found the Cronbach alpha coefficient was 0.909 for the Positive Affect Scale and 0.874 for the Negative Affect Scale.

\(^{11}\) Though our Trust game was understood as “one-shot” in its implementation, we expect that the evolved psychology applied in the game errs to caution by processing information about one-shot interactions with uncertain resource asymmetries under the premise that they may be repeated in the future (e.g., see Delton et al.,
3.3 Experimental procedures

The experiment was programmed using z-Tree (Fischbacher, 2007). There were eight experimental sessions, each lasting approximately thirty-five minutes. No participant participated more than once. Each session had between 18 and 24 participants, seated in individual cubicles, and was conducted as follows. An experimenter read the instructions aloud explaining experimental procedures and payoffs while every participant followed along with their own copy of the instructions. After finishing the instructions, participants were given five minutes to privately write down their answers to several quiz questions. After participants completed the quiz, the experimenter distributed a printed copy of the correct quiz answers. To ensure understanding, any remaining questions were answered privately.

Participants, randomly assigned to one of two roles: “person 1” (investor) or “person 2” (trustee), interacted anonymously in the Trust game over a local computer network, then completed the 20 item survey in which they reported the intensity of various emotional states consequent on their decisions, game interactions, and resulting outcomes. Earnings from the Trust game plus $7 for arriving to the experiment on time and participating were paid out privately at the end of the experiment.

2011). We also suspect that investors who make trust-based choices discover the consequent effects on their payoffs and extend this information when constructing generalizable models about the trustworthiness of trustees in the population (e.g., the experimental subject pool).
3.4 The Recalibrational model and Valence models

According to the Recalibrational model, post-interaction emotions are “triggered” once they integrate information about a Trust game outcome and computationally identify specific successes and failures. We label these computational triggers $S$ and $L$, for the short-sighted program’s goal achievement and the long-sighted program’s goal achievement, respectively.

We calculated success (with a maximum of 1 and minimum of 0) of the short-sighted program achieving its goal ($S$) according to competing perspectives of the investor (I) and trustee (T):

$$S = \begin{cases} \frac{e - s + r}{e + 2s} & \text{for I} \\ \frac{3s - r}{3s} & \text{for T when } s > 0 \\ 0 & \text{otherwise} \end{cases}$$

We calculated success (with a maximum of 1 and minimum of 0) of the long-sighted program achieving its goal ($L$), based on the mutual perspective shared by investor and trustee:

$$L = \text{Trust} \times \text{Trustworthiness},$$

where $\text{Trust} = s/e$, and $\text{Trustworthiness} = \min\{r/s, I\}$ if $s > 0$, else 0.

$S$ evaluates the short-sighted program’s goal achievement after investor and trustee decisions have been made. Both investor and trustee can maximize their short-term goal by keeping and not transferring available funds – decisions that result in large $S$ triggers. In addition to valuing any portion of the endowment kept, an investor’s short-sighted program values maximally recouping profitable returns on any investment made. Thus, to reasonably evaluate opportunity captured by an investor we consider how much of the endowment was
kept and how much of the multiplied investment was recouped by calculating \((e - s + r)/(e + 2s)\). Accordingly, an investor’s \(S\) trigger is maximized when all endowment was kept (in which case \(s = 0\) and \(r = 0\)), or if, in addition to any endowment kept, the maximum possible profitable return from the investment was recouped (i.e., \(r = 3s\)). An investor’s \(S\) is minimized after a Trust game in which a large amount was invested but nothing was returned. A trustee’s \(S\) is maximized after a game in which income was received and kept (i.e. where \(s > 0\) and \(r = 0\)) and is minimized after a game that generated no income (where \(s = 0\)) or after a game in which they returned everything and kept nothing (where \(s > 0\) and \(r = 3s\)).

The long-sighted program’s goal of consummating and maintaining a cooperative trust-based relationship requires that both trust and trustworthiness be demonstrated. Trust is demonstrated by the invested amount of endowment at risk. Trustworthiness is demonstrated by proportion of investment voluntarily reciprocated to the investor. As such, by making a larger transfer, one’s subsequent \(L\) trigger can increase, but is only maximized (mutually for investor and trustee) after a Trust game where investment was largest (\(s = e\)) and the investment was at least returned (\(r \geq s\)).

---

12 Our model explicitly uses \(s\) and \(r\) to compute functional outcomes triggering recalibrational emotional experience following trust based exchange. We would need to consider alternative computational forms if assuming social preferences such as inequity aversion (Fehr & Schmidt, 1999), “Rabin fairness” belief-based reciprocity (Rabin, 1993), or guilt-aversion (Battigalli & Dufwenberg, 2007). However, there is disagreement in the literature as to what a “correct” reference point is (e.g., Kahneman, 1992) and social preference theories lack convincing adaptive explanations for why such preferences should persist. Our Recalibrational theory proposes that trust-based exchange behavior can be understood in terms of tradeoffs between adaptive short-sighted and long-sighted goals for which relative levels of investment and return on investment are the most fundamental reference points upon which to recalibrate future behavioral propensities.

13 Prior research has found that when investment (\(s\)) is relatively large (e.g., greater than half of the endowment), \(r\) tends to exceed \(s\), whereas when \(s\) is relatively small, \(r\) tends to be equal or less than \(s\) (e.g., see Ostrom & Walker, 2003). Given we find this same distribution, the above predictions should hold: as \(s\) and \(r\) increase in game interactions, the resulting \(L\) becomes larger.
established when either the investor or the trustee has pursued maximum opportunism. As such, \( L = 0 \) when \( s = 0 \) or when \( r = 0 \).

The relationship between \( s \) and \( r \) should also be predictive of emotion activation. Notice that \( L \) increases with \( s \). For a fixed \( s \), \( L \) also increases with \( r \). The investor’s \( S \) decreases with \( s \), and increases with \( r \) for a fixed \( s \), while the trustee’s \( S \) increases with \( s \) and decreases with \( r \) for a fixed \( s \). \( L \) affects Sets 1, 2, 3, and 4, while \( S \) affect Sets 1 and 3. Let us hold \( s \) constant and assume the loadings on \( L \) and \( S \) are approximately equal. If we compare scenarios where \( r < s \) to scenarios where \( r > s \), the aforementioned comparative statistics would predict that (i) there is higher emotion activation when \( r > s \), and (ii) there is more conflicted emotion activation when \( s > r \). The first prediction is due to up-regulation of Sets 1 and 2 and down-regulation of Sets 3 and 4. The second prediction is due to down-regulation of all sets, creating “conflicted” emotion activation of both negative and positive sets, albeit lesser positive emotion activation than where \( r > s \). We can also see conflicted emotion activation prediction by comparing the relative values of \( L \) and \( S \), as we expect to see conflicted emotions when the values depart.

When \( r < s \) then \( L = \frac{s}{e} \times \frac{r}{s^2} \), but when \( r \geq s \) then \( L = \frac{s}{e} \). Comparing the difference between \( L \) and \( S \) for all outcomes when \( r < s \) versus when \( r \geq s \), the difference is larger for both the investor and trustee when \( r < s \).\(^{14}\)

\(^{14}\) For the investor the difference between \( S \) and \( L \) is \( \int_0^e \int_0^3 \left( \frac{e^{-s} - r}{e + 2s} - \frac{s}{e} \right) \, dr \, ds \) - \( \int_0^e \int_0^{3s} \left( \frac{e^{-s} - r}{e + 2s} - \frac{s}{e} \right) \, dr \, ds > 0 \) using \( e = 10 \). Note that we limited the amount of the trustee return \( r \) to half the total trustee receives so that we are only integrating over outcomes observed in prior experiments. For the trustee the difference is \( \int_0^e \int_0^{3s} \left( \frac{3s - r}{3s} - \frac{s}{e} \right) \, dr \, ds > 0 \).
In summary, the Recalibrational model expects emotions to be triggered by $L$ and $S$: computational assessments of short- and long-sighted programs’ successes and failures. Positive emotions are maximally experienced when trigger values are largest and negative emotions are maximally experienced when trigger values are smallest. According to their design functions, triggered emotions either contribute to the reinforcement of successes or the reduction of failures by up-regulating or down-regulating specific programs in self and others. We tested whether constellations of specific antecedents (the $L$ and $S$ triggers produced by Trust game interaction) reliably predict specific sets of emotional experiences.

We applied confirmatory factor analysis (CFA) to emotional reports to verify the four factor structure interpreted by the Recalibrational model and subsequently used in our structural equation model (SEM). We also used CFA to measure how well the reported emotional states fit the three variants of the Valence model. From these CFA results of basic Valence models we chose the best fit model to compare for fit with the Recalibrational model.

We applied SEM to compare the fit of the Recalibrational model and the Valence model to observed data. Schematics of our SEM models are provided in the results section below. Tests of SEMs allow us to answer how well the conceptual models of interest fit the data, and whether the model we are suggesting shows better fit than rival models. Our results provide evidence that a multivariate Recalibrational model significantly outperforms the Valence models when describing the patterned experience of emotions reported after a Trust game. These results support the hypothesis that sets of recalibrational emotions are triggered in patterned response to the adaptive problems produced by trust-based interactions.
4. Results

In this section, we first report general results of the Trust game and the emotional status survey. Next, we investigate whether the experience of 20 emotions conforms to predictions of the Recalibrational model or predictions of Valence models. Finally, we examine the full models of emotional experiences, comparing the fit of the unrestricted Valence model and the Recalibrational model.

We found no significant differences between genders or between the eight sessions and report the joint results of all 170 participants. Figure 1 displays the scatter plot of the amount sent and the amount returned in Trust games. There was substantial variability in individual behavior. On average, investors sent $6.01 (SD = 3.64) and trustees returned $6.16 (SD = 5.92), resulting in profits of $10.14 (SD = 3.72) and $11.88 (SD = 7.12), respectively. These results are consistent with previous findings of Berg et al. (1995). Likewise, there was substantial variability in individual reports of emotional experience. The average reported emotional state (as a result of Trust game interactions) had a mean of 2.20 (median = 1, SD = 1.45), near 2 (“a little”). Ratings on every emotional state ranged from 1 (“very slightly or not at all”) to 5 (“extremely”). While the modal report for most (17/20) emotional states was 1 (“very slightly or not at all”), modes were also at 3 for happy and 5 for content and appreciative. Reports of 1 were more frequent for emotional states in the negative set than for the positive set (1218/1700 versus 527/1700, respectively), contributing to significantly lower intensity of reported negative states (M = 1.61, SD = .77) than positive states (M = 2.80, SD = 1.08) according to Wilcoxon matched-pairs tests (Z = 7.605, p < .001). This pattern of
Conflicted Emotions Following Trust-based Interaction

significantly lower reported negative states was observed in both investors \((Z = 5.853, p < .001)\) and trustees \((Z = 4.888, p < .001)\).

We constructed an “activation” score based on individuals’ average reports across all emotions. Where trust and trustworthiness was demonstrated \((s < r)\), both investors and trustees experienced more emotional activation \((\text{investor Mdn} = 2.40, \text{trustee Mdn} = 2.55, 34 \text{ pairs})\) than where trustworthiness was not demonstrated \((\text{investor Mdn} = 2.08, \text{trustee Mdn} = 2.12, 40 \text{ pairs})\), a significant difference according to the Wilcoxon rank-sum (Mann-Whitney) test \((\text{investor: } Z = 2.834, p = .005; \text{trustee: } Z = 3.821, p < .001)\). As shown in Table 2, this effect of higher emotional activation where trustworthiness is demonstrated is driven by the positive emotions of sets 1 and 2, which were compared to median emotional activation levels of negative emotions using a Wilcoxon rank-sum (Mann-Whitney) test \((\text{investor: } Z = 5.605, p < .001; \text{trustee: } Z = 6.262, p < .001)\). Furthermore, the Wilcoxon rank sum test \((\text{investor: } Z = 3.249, p = .0012; \text{trustee: } Z = 5.695, p < .001)\) shows that the emotional activation effect after demonstrated trustworthiness was robust among the long-sighted emotions exclusive to set 2. The social emotions of Set 2 were more activated when trustworthiness was demonstrated \((\text{investor M}=3.13, \text{SEM}=0.17; \text{trustee M}=3.66, \text{SEM}=0.18)\) than when trustworthiness was not demonstrated \((\text{investor M}=2.32, \text{SEM}=0.14; \text{trustee M}=1.96, \text{SEM}=0.15)\).

As shown in Table 2, a Wilcoxon rank-sum (Mann-Whitney) test \((\text{investor: } Z = 4.987, p < .001; \text{trustee: } Z = 5.758 p < .001)\) indicated that a ‘broken trust’ effect also exists: where trustworthiness was not demonstrated, both investors and trustees reported higher activation of
negative emotions than where it was. Furthermore, a Wilcoxon rank-sum test (investor: $Z = 3.923, p < .001$; trustee: $Z = 2.149, p = .032$) shows that this ‘broken trust’ effect was robust among the negative long-sighted emotions exclusive to Set 4. The negative social emotions of Set 4 were more activated when trustworthiness was not demonstrated (investor $M=1.66$, SEM=0.14; trustee $M=1.79$, SEM=0.16) than when it was (investor $M=1.07$, SEM=0.04; trustee $M=1.31$, SEM=0.11).

### 4.1 Shared features of emotions

Valence models assume two factors: one comprised of a standard set of positive emotional states that positively correlate with one another [appreciative, happy, content, cheerful, triumphant, inspired, secure, proud, believable, and surprised], and the other comprised of a standard set of negative emotional states that positively correlate with one another [disgusted, jealous, aggravated, frustrated, angry, depressed, sad, embarrassed, ashamed, and guilty]. Item analysis indicated that not all (43 of 45) correlations were significantly positive between positive states, nor between all (36 of 45) negative states.

Consistent with P3 and P4, we observed occurrences of simultaneously activated positive and negative emotions. Investors reported experiencing simultaneous activation of emotions in Set 4 (especially guiltiness) and Set 1 (especially contentment) while trustees reported simultaneous activation of emotions in Set 3 (especially anger) and Set 2 (especially pride).

From cross tabulation, we observed 57 cases from 13 (7.64% of) respondents reporting activation of positively valenced (P) and negatively valenced (N) emotions that were both felt “extremely” (= 5); 231 cases from 34 (20% of) respondents reporting activation of P and N
emotions that were both felt in the range from “quite a bit” to “extremely” (≥ 4); 973 cases from 69 (40.59% of) respondents reporting activation of P and N emotions that were both felt in the range from “moderately” to “extremely” (≥ 3); and 2,653 cases from 114 (67.06%) respondents reporting positive and negative states that were both felt in the range from “a little” to “extremely” (≥ 2). We also examined simultaneous activation of the 9 positively valenced (P) emotions (not including surprise) and the 10 negatively valenced (N) emotions and constructed a “conflicted” score (valued 1-5) based on the maximum level at which any pair of P and N emotions were both equally activated (i.e., the min{max P, max N}). We found that where trust was extended but trustworthiness not demonstrated (0 < s and s ≥ r), both investors and trustees experienced more conflicted emotions (investor Mdn =2.00, trustee Mdn =3.00, 40 pairs) than did the investors and trustees (investor Mdn =1.00, trustee Mdn =1.00, N=34 pairs) where trustworthiness was demonstrated (0 < s < r).

4.2 Comparison of Confirmatory Factor Analysis fit

We used the Stata v.12 software’s Confirmatory Factor Analysis (CFA) to measure how well the reported emotional states fit the three variants of the Valence model and the Recalibrational model. Each variant of the Valence model shares the assumption that positive correlations exist among individuals’ reported positive states and positive correlations exist among individuals’ reported negative states. Therefore, in all Valence models we constrained

15 We used Stata to fit CFA with maximum likelihood method using Stata’s modified Newton-Raphson optimization algorithm. Tolerances for convergence were $e^{-6}$ for the coefficient vector, $e^{-7}$ for log likelihood, and $e^{-5}$ for the scaled gradient. All models converged before reaching the maximum amount of iterations. Standard errors calculated using the observed information matrix (OIM).
Conflicted Emotions Following Trust-based Interaction

each emotion to load onto only one of the two factors. However, because each variant of the
Valence model has a different assumption concerning relationships that might exist among
simultaneously experienced positive and negative emotions, they differ only in the constraints
imposed on the positive and negative factor correlations. Following the prediction P1, the
uncorrelated Valence model 1 constrains the factors to have a zero correlation (where positive
emotions bear no relationship with negative emotions). Following P2, the correlated Valence
model 2 constrains the factors to a correlation of negative one (as would be appropriate if the
experience of emotions was only possible on a bipolar valence continuum). Following P3, the
unrestricted Valence model 3 imposes no restrictions on the factors’ correlation.

Summaries of CFA results for the three Valence models and the Recalibrational model are
shown in Table 3. The lesser Bayesian information criterion (BIC), lesser root mean square
error of approximation (RMSEA), greater comparative fit index (CFI), and greater log-
likelihood (LL) made it apparent that the correlate Valence model fits better than the
uncorrelated Valence model, and that the unrestricted Valence model fits better than both
correlated and uncorrelated Valence models. The difference between the unrestricted Valence
model and the uncorrelated Valence model was statistically significant ($\chi^2(1) = 100.90$, $p <
.001$). Consistent with P3, the unrestricted Valence model’s correlation between positive and
negative factors was $-.70$ (95% CI [-.787, -.614]) and significantly different from zero and
from -1 at a 1% level of significance. In Table 4 we describe the derived CFA fit from the
Recalibrational model. The Recalibrational model predicted the patterned experience of
emotions according to the four factors corresponding to Set 1, Set 2, Set 3, and Set 4 (see Table
1). Consistent with P4, all emotional states loaded positively and significantly (at a 1% level) onto the predicted latent factors of the Recalibrational model, but not the predicted latent factors of the unrestricted Valance model. With a greater LL, greater CFI, lesser RMSEA and lesser BIC, the Recalibrational model provides a better fit than the unrestricted Valence model (according to guidelines set forth by Gefen et al., 2011). This provides support for our final prediction P5.

4.3 Comparison of structural equation model fit

We used the Stata v.12 software’s Structural Equation Modeling (SEM) to compare the fit of the Recalibrational model to the fit of the unrestricted Valence model. To compare these models we tested both with variables S and L, as computed from game interactions. Diagrams of these SEMs are provided in Figures 2 and 3 and results provided in Figures 2 and 3 and Table 4. Given that previously we did not find support for either perfect independence or interdependence, we did not restrict correlations and allowed all latent factors to freely correlate in both models. The Recalibrational model has more factors than the Valence model, which can arguably lead to “overfitting” – having a better fit by describing more error instead of predicted relationship. To avoid overfitting, we report the BIC, which penalizes for added

\[ 16 \] We used Stata to fit SEM with maximum likelihood method using Stata's modified Newton-Raphson optimization algorithm. Tolerances for convergence were $e^{-6}$ for the coefficient vector, $e^{-7}$ for log likelihood, and $e^{-5}$ for the scaled gradient. All models converged before reaching the maximum amount of iterations. Standard errors calculated using the observed information matrix (OIM). The models and standardized estimates are illustrated in Figures 2 and 3.
variables. Finally, we report the difference between models, assessing whether the better fit was statistically significant despite the difference in factors seen in Table 4.

As with CFA, we found superior results via SEM for the Recalibrational model, consistent with P5. Despite penalizing for additional fitted variables, the difference was significant. In the Recalibration model, all of the emotions loaded onto latent factors with significant coefficients (below 5%) and predicted signs (see Figure 3), whereas in the unrestricted Valence model not all (17 of 19) of the tested emotions loaded onto latent factors with significant coefficients (see Figure 2). In particular, guilt (involved with contentment in conflicted response to broken trust) and shame were not predicted by the Valence model.

5. Discussion

Using CFA to assess latent sets, and SEMs to assess Trust game triggers on latent sets we demonstrated that the Recalibrational model predicts the experience of four latent sets of emotions following the Trust game, strongly and significantly outperforming the Valence models. Unlike simpler Valence models, the Recalibration model predicts activation of conflicted emotions in social dilemmas, a topic that has received relatively little attention. In addition to better fit, our Recalibrational model is interpretable: it is derived from functional accounts of evolved emotions that respond to adaptive problems we believe to be produced by trust-based cooperative dilemmas such as the sequential choice Trust game and the
Conflicted Emotions Following Trust-based Interaction

simultaneous choice Prisoner’s Dilemma. To provide additional support for the Recalibrational perspective, it will be helpful to have tests of whether the conflicted emotions – such as evoked in Trust games – are only triggered by situations with social context (where a long-sighted program would be concerned) or equally by risk game situations in a non-social context (where computers are randomly making decisions).\(^{17}\) We suspect that in a non-social context, the individuals who decide to keep their endowment rather than gamble it will not feel the same “guilty contentment” that non-trusting investors report, likewise the gambler who incurs a loss on a wager will not experience the “angry pride” otherwise felt by the investor who invested much of their endowment only to have their trust broken with a small return. Future research could adapt emotion measures like ours to social vs non-social treatment designs similar to those used by Houser, Schunk, & Winter (2010) or Chang et al. (2010) to evaluate whether conflicted emotions are equally triggered in non-social situations.

Below we discuss potential sources of unexplained variance limiting our Recalibrational model’s explanatory ability and how future research might deal with these limitations. Finally, we conclude discussing the theoretical implications of having conflicted goals and the economic implications of having conflicted emotions.

Unexplained variance in how strongly emotional experiences are rated by participants may be due to imperfect awareness of one’s own emotions, differing interpretations of the emotion labels, or reports with compromised fidelity due to noise or dishonesty. People who are asked to rate single emotions may not be able to accurately describe their emotions (Ellsworth &

\(^{17}\) We thank an anonymous reviewer for bringing this suggestion to our attention.
Conflicted Emotions Following Trust-based Interaction

Tong, 2006) if emotion experiences are more often and accurately described with multiple words (Izard, 1977), or with different words among different people. While we acknowledge that language could present problems for this research, the success of previous research on self-reported emotions in conjunction with experimental games (e.g., Ketelaar & Au, 2003; Kausel & Connolly, 2014) gave us encouragement in pursuing measures of self-reported emotions following an economic game.18

Data quality could also have been affected if participants made untruthful reports. Experimental economists are particularly concerned that participants “will not ‘tell the truth’ unless incentives make truth telling compatible with maximizing utility” (Lopes, 1994, p.218). According to a meta-review by Camerer and Hogarth (1999) there is no clear evidence that additional financial incentives would improve the quality of responses in a simple survey task like ours. In fact, it has been noted that for short tasks like PANAS surveys that people are known to voluntarily complete without problem, an attempt at increasing participation via financial incentives often “backfires” with counter-intentional effects (e.g., see Mellstrom & Johannesson, 2008). Nevertheless, wary of the possibility that participants may have been incentivized to use efficiency tactics to expediently complete the survey, such as by marking all responses the same, we reviewed our data and found only one apparent case of such behavior (< 1% of sample).19

18 While our study proceeded without problem using English language tokens as proxies of activated evolved functional systems, we expect a similar approach could be applied cross-culturally: working with the hypothesis that evolved functional design features exist among sets of experienced emotions, researchers knowledgeable of other cultures and languages could classify existing language tokens of emotional states into functionally distinct categories and investigated their activation and effects accordingly.

19 One individual reported 3 on all emotions.
Conflicted emotions have been noted in many economic and organizational contexts, yet it remains unclear what the economic consequences of mixed emotions might be for marketing and industrial organization. The use of advertisement and marketing strategies arousing conflicted emotions may be effective (e.g., for individuals more accepting of paradox and duality, Williams & Aaker, 2002) or ineffective (Hong & Lee, 2010), for example, because conflicted emotions make people feel “torn” and uncomfortable (Williams & Aaker, 2002) or make ads less memorable (Aaker, Drolet, & Griffin, 2008). In organizational contexts, conflicted emotions are regularly encountered among employees and managers (Pratt, 2000; Pratt & Doucet, 2000; Fong & Tiedens, 2002; Sellers, 2003). Conflicted emotions could harm employee or managerial performance once expressed (e.g., by provoking dominating behavior by observers, Rothman, 2011), or could improve performance by enhancing resilience and ability to cope with stressful events (Larsen, Hemenover, Norris, & Cacioppo 2003), by increasing creativity (Fong, 2006), or improving judgment accuracy (Reese et al. 2013). The experience of conflicted emotions among experts (e.g. doctors) or leaders may also be undesired by laypeople (e.g. patients) who defer to experts and leaders for categorizations and decisions (e.g. “which disease do I have and which treatment do I need?”) and expect certainty and decisiveness of them, not ambivalence or conflict (Marsh & Rothman, 2013). Likewise, conflicted emotions triggered after interactions where trust was demonstrated but trustworthiness was not could lead to future distrust. More research is needed to uncover the consequences of conflicted emotions in the context of trust-based interactions.

While some researchers have moved past the bipolar affect models, instead recognizing that positive and negative affect are at times independent dimensions, psychophysiologists
(Lang et al., 1993; Driscoll, Tranel, & Anderson, 2009) neuroscientists (Proverbio, Zani, & Adorni, 2008; Screenivas, Boehm, & Linden, 2012; Xiang, Lohrenz, & Montague, 2013) behavioral economists (Morretti & di Pellegrino, 2010; Brandts, Riedl, & van Winden, 2009; Van den Berg, Dewitte, & Warlop, 2008; Morris, 1995) and decision scientists (Hogarth, Portell, Cuxart, & Kolev, 2011; Reid & Gonzalez-Vallejo, 2009; Schlosser, Dunning, & Fetchenhauer, 2013) continue to use bipolar affect scales, whether based on versions of the PANAS or the Self-Assessment-Manikin valence scale developed by Lang (1980). Our study cautions against assuming that the explanatory power provided by the Valence model is sufficient for understanding relationships between trust-based behavior and emotions. We suggest that more complex multivariate models, such as derived from a Recalibrational perspective, better explain the triggered experience of conflicted emotions and subsequent behaviors.

Our study has demonstrated that an evolutionary–functional framework is a productive and promising approach to uncovering the situations following trust-based interactions that trigger recalibrational emotions and conflicted feelings. Contrary to popular opinion, these findings suggest that the experience of conflicted emotions, such as following certain trust-based interactions, is normal.

**Acknowledgements:** We thank Larry Fiddick, Joaquín Gómez-Miñambres, Daniel Nettle, Nat Wilcox, and participants at the 2012 Human Behavior and Evolution Society meetings for helpful feedback. We are grateful to Chapman University and the Economic Science Institute for research support.
6. References


Conflicted Emotions Following Trust-based Interaction


Conflicted Emotions Following Trust-based Interaction


Conflicted Emotions Following Trust-based Interaction


Conflicted Emotions Following Trust-based Interaction


Conflicted Emotions Following Trust-based Interaction


Conflicted Emotions Following Trust-based Interaction


### Table 1: Functional taxonomy of emotional states

<table>
<thead>
<tr>
<th>Set</th>
<th>Emotional State</th>
<th>Functional Features</th>
<th>Facilitating Adaptive Goal(s)</th>
<th>Recalibrational Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>U: Long-sighted</td>
<td>V: Short-sighted</td>
</tr>
<tr>
<td>1</td>
<td>Appreciative</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Content</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Cheerful</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Triumphant</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Inspired</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Secure</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>2</td>
<td>Proud</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Believable</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Disgusted</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Jealous</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Aggravated</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td>Frustrated</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Angry</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Depressed</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4</td>
<td>Embarrassed</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Ashamed</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>Guilty</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>5</td>
<td>Surprised</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Note: X’s indicate that the emotional states have functions corresponding to those columns.
### Table 2: Mean positive and negative emotional activation after broken trust or demonstrated trustworthiness

<table>
<thead>
<tr>
<th></th>
<th>Investors</th>
<th></th>
<th>Trustee</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Broken</td>
<td>Demonstrated</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trust</td>
<td>Trustworthiness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average of:</td>
<td>N = 51</td>
<td>N = 34</td>
<td>N = 51</td>
<td>N = 34</td>
</tr>
<tr>
<td>Recalibrational Set 1</td>
<td>2.22 ***</td>
<td>3.65</td>
<td>2.19 ***</td>
<td>3.91</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.15)</td>
<td>(0.15)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>Recalibrational Set 2</td>
<td>2.32 **</td>
<td>3.13</td>
<td>1.96 ***</td>
<td>3.66</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(.017)</td>
<td>(0.15)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>All Positive Emotions</td>
<td>2.25 ***</td>
<td>3.54</td>
<td>2.14 ***</td>
<td>3.85</td>
</tr>
<tr>
<td></td>
<td>(0.11)</td>
<td>(0.14)</td>
<td>(0.14)</td>
<td>(0.15)</td>
</tr>
<tr>
<td>Recalibrational Set 3</td>
<td>1.87 ***</td>
<td>1.13</td>
<td>2.13 ***</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.05)</td>
<td>(0.15)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Recalibrational Set 4</td>
<td>1.66 ***</td>
<td>1.07</td>
<td>1.79 *</td>
<td>1.31</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.04)</td>
<td>(0.16)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>All Negative Emotions</td>
<td>1.81 ***</td>
<td>1.11</td>
<td>2.03 ***</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>(0.12)</td>
<td>(0.04)</td>
<td>(0.12)</td>
<td>(0.06)</td>
</tr>
</tbody>
</table>

Note: ***, **, and * denote value different from the average trustworthy case at p < .001, .01, and .05, respectively, according to the Wilcoxon test. The average reported value (on a scale 1-5) of “All Positive Emotions” (Recalibrational sets #1 and #2) and “All Negative Emotions” (Recalibrational sets #3 and #4), with standard error in parenthesis, is given for cases with broken trust (returned not greater than positive amount sent) and for cases where trustworthiness was demonstrated (i.e., amount returned greater than amount sent).
### Table 3: Details of Confirmatory Factor Analyses

<table>
<thead>
<tr>
<th>Model</th>
<th>Specification</th>
<th>DF</th>
<th>Log-likelihood (LL)</th>
<th>Root Mean Square Error of Approximation (RMSEA) [90% CI]</th>
<th>Comparative Fit Index (CFI)</th>
<th>Bayesian Information Criteria (BIC)</th>
<th>Difference in fit compared to model R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Independent Valence model: Zero (0) PA/NA c orrelation</td>
<td>57</td>
<td>-4,440.83</td>
<td>.145 [.134, .156]</td>
<td>.767</td>
<td>9,174</td>
<td>$\chi^2(6)= 148.39$ p &lt; .001</td>
</tr>
<tr>
<td></td>
<td>Bipolar Valence model: Negative (-1) PA/NA c orrelation</td>
<td>57</td>
<td>-4,422.53</td>
<td>.140 [.129, .151]</td>
<td>.783</td>
<td>9,138</td>
<td>$\chi^2(6)= 130.09$ p &lt; .001</td>
</tr>
<tr>
<td></td>
<td>Unrestricted Valence model: Unrestricted PA/NA c orrelation</td>
<td>58</td>
<td>-4,393.34</td>
<td>.132 [.121, .143]</td>
<td>.808</td>
<td>9,085</td>
<td>$\chi^2(5)= 100.90$ p &lt; .001</td>
</tr>
<tr>
<td>R</td>
<td>Recalibrational model: 4 factors</td>
<td>63</td>
<td>-4,292.44</td>
<td>.100 [.089, .112]</td>
<td>.892</td>
<td>8,908</td>
<td></td>
</tr>
</tbody>
</table>
Conflicted Emotions Following Trust-based Interaction

Table 4: Details of structural equation fit analyses

<table>
<thead>
<tr>
<th>Model</th>
<th>DF</th>
<th>Log-likelihood (LL)</th>
<th>Root Mean Square Error of Approximation (RMSEA) [90% CI]</th>
<th>Comparative Fit Index (CFI)</th>
<th>Bayesian Information Criteria (BIC)</th>
<th>Difference in fit compared to Recalibrational Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unrestricted Valence</td>
<td>62</td>
<td>-4,809.33</td>
<td>.126 [.116, .136]</td>
<td>.799</td>
<td>9.937</td>
<td>$\chi^2(7) = 104.77$ p &lt; .001</td>
</tr>
<tr>
<td>Recalibrational</td>
<td>69</td>
<td>-4,704.56</td>
<td>.099 [.088, .109]</td>
<td>.880</td>
<td>9.763</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Bubble plot of the amount sent and the amount returned

Note: Observations were plotted with bubbles, where the relative size indicates the number of observations. The smallest bubble plotted represents one observation and the largest bubble plotted represents eight observations. The red colored bubbles are observations (where $0 < \text{sent} \geq \text{returned}$) that are significantly more “conflicted” (*** denotes $p < .001$) than among all other observations according to a Wilcoxon rank-sum (Mann-Whitney) test.
Conflicted Emotions Following Trust-based Interaction

Figure 2: “Valence” SEM of emotion activation following a Trust Game

Note: Nineteen different emotional experiences reported by participants are indicators of two latent sets (PA and NA) of the Valence model. S and L triggers, based on individuals’ Trust game decisions, load onto these latent factors. Standardized coefficient (standard error) reported on path. ** indicates $p \leq .01$, *** indicates $p \leq .001$. 
Figure 3: “Recalibrational” SEM of emotion activation following a Trust Game

Note: Nineteen different emotional experiences reported by participants are indicators of four latent sets (Sets 1, 2, 3, 4) of the Recalibration model. S and L triggers, based on individuals’ Trust game decisions, load onto these latent factors. Standardized coefficient (standard error) reported on path. ** indicates $p \leq .01$, *** indicates $p \leq .001$. 
Appendix A – Instructions

This is an experiment in the economics of decision-making. Various research agencies have provided funds for this research. The currency used in the experiment is experimental dollars, and they will be converted to U.S. Dollars at a rate of 1 experimental dollar to 1 dollar. At the end of the experiment your earnings will be paid to you in private and in cash. It is very important that you remain silent and do not look at other people's work. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc. you will be asked to leave and you will not be paid. We expect, and appreciate, you adhering to these policies.

THE EXPERIMENT

The participants in today's experiment will be randomly assigned into two-person groups. In addition to the group assignment each participant will also be randomly assigned to a specific type in the group, designated as Person 1 or Person 2. You and the other participant in your group will make choices that will determine your payoffs. The experiment consists of two decision stages.

In stage 1, Person 1 receives $10 and then decides how many dollars to send to Person 2. Person 1 can send none, more than none, or all of the $10 to Person 2. The amount sent by Person 1 is tripled before reaching Person 2. In stage 2, Person 2 decides how many of the dollars they received to send back to Person 1. Person 2 can send none, more than none, or all of the amount received back to Person 1. At that point the experiment is over.

Next we describe in details the decisions made by both persons in each stage of the experiment. **Stage 1:** Person 1 receives $10 and then decides how many dollars to send to Person 2. Person 1 can send none, more than none, or all of the $10 to Person 2. The amount sent by Person 1 in the box labeled “The amount sent by Person 1” below. Person 1 keeps any amount that is not sent to Person 2. The amount sent by Person 1 is tripled before reaching Person 2.
**Stage 2:** After learning the amount sent by Person 1, Person 2 decides how many dollars to send back to Person 1. Person 2 can send none, more than none, or all of the amount in Person 2’s account at that time. Person 2 enters the amount sent back to Person 1 in the box labeled “The amount sent back by Person 2” below. The amount sent back by Person 2 is NOT multiplied. Person 2 keeps any amount that is not sent back to Person 1.

Finally, at the end of the Stage 2 the total earnings are reported to each person.

- Person 1's earnings will equal $10 less the amount sent to Person 2 plus the amount sent back by Person 2.
- Person 2's earning will equal three times the amount sent by Person 1 less the amount sent back to Person 1.

Please record the decisions and your earnings on your record sheet under the appropriate heading.

**SUMMARY**

The computer will assign you and one other participant to a two-person group, consisting of **Person 1** and **Person 2**. In stage 1, Person 1 receives $10 and then decides how many dollars to send to Person 2. Person 1 can send none, more than none, or all of the $10. The amount sent by Person 1 is tripled. In stage 2, Person 2 decides how many dollars to send back to Person 1. Person 2 can send none, more than none, or all of the amount in Person 2's account at that time. At the end of Stage 2 the total earnings are reported to each person. This experiment is now over and your earnings will be part of the total you will be paid.
QUIZ

Before starting, we want you to answer some questions regarding the experiment to be sure you understand what will follow. After five minutes an experimenter will return to privately review your answers. Afterwards you will participate in the experiment only one time.

- True or false: the amount sent by Person 1 is tripled before reaching Person 2's account.
- True or false: the amount sent back by Person 2 is tripled before reaching Person 1's account.
- What is the largest amount Person 1 can send to Person 2?
- What is the smallest amount Person 2 can send back to Person 1?
- If Person 1 sent $4.20 to Person 2, what is largest amount Person 2 can send back to Person 1?
- If Person 1 sent $9.00 to Person 2, what is smallest amount Person 2 can send back to Person 1?
- True or false: If Person 1 sends something to Person 2, then Person 2 has to send something back to Person 1.
- True or false: you will participate in this experiment only one time.
Appendix B – 20-item Emotion Survey

The following scale consists of a number of words that describe different feelings and emotions. Read each item and then mark the appropriate answer in the space next to that word. Indicate how well the experience in which you just participated made you feel.

Use the following scale to record your answers:
(1) very slightly or not at all, (2) a little, (3) moderately, (4) quite a bit, (5) extremely

<table>
<thead>
<tr>
<th>Guilty</th>
<th>Secure</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Embarrassed</td>
<td>Angry</td>
<td></td>
</tr>
<tr>
<td>Proud</td>
<td>Disgusted</td>
<td></td>
</tr>
<tr>
<td>Ashamed</td>
<td>Jealous</td>
<td></td>
</tr>
<tr>
<td>Inspired</td>
<td>Surprised</td>
<td></td>
</tr>
<tr>
<td>Depressed</td>
<td>Appreciative</td>
<td></td>
</tr>
<tr>
<td>Believable</td>
<td>Cheerful</td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>Aggravated</td>
<td></td>
</tr>
<tr>
<td>Happy</td>
<td>Frustrated</td>
<td></td>
</tr>
<tr>
<td>Triumphant</td>
<td>Sad</td>
<td></td>
</tr>
</tbody>
</table>

OR