Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Behavioral Sciences

Social emotions and psychological games

Luke J Chang[1,3] and Alec Smith[2,3]

Emotions arise from cognitive appraisals and organize adaptive behavioral responses. The appraisals associated with social emotions such as guilt and anger can be modeled with utility functions that depend on both material and psychological payoffs, and their effect on behavior can be mathematically described using game theory. Guilt arises from the belief that an agent has disappointed a relationship partner and motivates reparative actions, while anger arises from the frustration of a goal being unexpectedly blocked and motivates aggressive actions. These psychological payoffs not only enable cooperation, but also appear to be associated with neural activations consistent with negative affective states. We believe integrating appraisal theory with game theoretic modeling can improve our ability to study emotions and predict behavior in social interactions.

Addresses
[1] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA
[2] Department of Economics (MC 0316), Virginia Tech, 880 West Campus Drive, Blacksburg, VA 24061, USA

Corresponding authors: Chang, Luke J (luke.j.chang@dartmouth.edu) and Smith, Alec (alecsmith@vt.edu)

[3] Both authors contributed equally to this work, and author order is alphabetical.

Current Opinion in Behavioral Sciences 2015, **5**:133–140

This review comes from a themed issue on **Neuroeconomics**

Edited by **John P O'Doherty** and **Colin C Camerer**

For a complete overview see the Issue and the Editorial

Available online 24th October 2015

http://dx.doi.org/10.1016/j.cobeha.2015.09.010

We spend much of our waking day engaged in collaborative social exchanges. Social emotions, in particular, are central in ensuring the success of these interactions [1]. For example, consider a scientific collaboration in which A shares data with B to perform a specialized analytic technique with the goal of publishing the results together. B would likely feel guilty if he shirked his responsibility and in response A would be angry that she wasted her time trusting B. Thus, social emotions play a critical role in ensuring a successful exchange. These emotions can be modeled using utility functions that incorporate both material and psychological payoffs, and their effect on behavior can be mathematically described using game theory. In this article we focus on two social emotions, guilt and anger, and demonstrate how game-theoretic models of these emotions capture important aspects of social behavior.

Psychological models of emotion

Emotions are psychological states comprised of multiple interrelated processes such as cognitive appraisals, physiological responses, behavioral action tendencies, and the phenomenological experience of feelings. Though there are many different perspectives on emotion ranging from categorical models of discrete emotions [2,3], multi-dimensional factor models [4,5], and psychological constructionist models [6,7], none have been more amenable to computational modeling than the cognitive framework of appraisal theory [8–11]. Appraisal theory defines emotions as adaptive responses that are elicited based on how an agent evaluates its situation (e.g., novelty, valence, threat, contamination, social norms, among others) [9•,12,13,14•,15]. Appraisals are typically directly related to the motivational goals of the agent (e.g., basic needs, safety, cultural values, beliefs) and occur in response to both external stimuli and also to internally generated thoughts, for example, when the agent is imagining the future or remembering the past. Agents continually interpret their environment with respect to their motivational goals and these evaluations or appraisals give rise to different feeling states that evolve as information changes [9•,14•]. Appraisals are thus cognitive antecedents to the experience of the emotion, though it remains an open question whether emotions are a consequence of appraisals or if the appraisal itself constitutes the emotional experience [16••].

In our view, appraisals precede emotions, which in turn prepare the agent to make adaptive responses via action tendencies [17]. Action readiness is the state of translating feelings and goals into behavioral actions. These actions could be as simple as approaching or avoiding a stimulus [18], or could take the form of embodied action preparations [19,20]. Whereas appraisals describe the inputs of the emotional experience, action tendencies delineate the behavioral outputs. This input–output view of emotion provides a structure that can be translated into mathematical models.

In this paper we focus on guilt and anger, two emotions that arise from social interactions and which can be described in terms of cognitive appraisals and action tendencies [21•].[4] For example, guilt arises from the appraisal that one has failed to live up to the expectations of a relationship partner [24••] and motivates reparative action tendencies [25–27]. Anger, in contrast, arises from the appraisal that progress toward a goal is blocked, or a social/moral norm has been transgressed [28•,29], and motivates punishment and revenge action tendencies [16••,30]. Using a theoretical approach known as psychological game theory [31•,32••] the appraisals associated with these emotions may be captured as the changes in an agent's expected payoff following a new event or outcome. These belief dependent appraisals can then be directly incorporated into the agent's utility function as psychological payoffs (i.e., subjective feelings) to capture the action tendencies associated with emotions.

Game theoretic models of emotion

Game theory is a set of mathematical tools for modeling interactive decision-making. These include mathematical descriptions of the strategies available to the players and of the payoffs (or utilities) resulting from those strategies. Additional details may include the sequence of play, the actions available to each player at each stage of the game, and the information available to each player. Players' beliefs are represented via probability distributions over actions, states, or other players' beliefs [33]. When combined with solution concepts such as Nash equilibrium [34], subgame perfect equilibrium [35], or sequential equilibrium [36], the formal structure of game theory provides predictions about how the game will be played and the payoffs to each player.

Early game theoretic models and applications assumed that agents behaved selfishly in maximizing their material self-interest: that is, each player's utility function depended only upon his own payoff. These models of purely self-interested individuals perform poorly in predicting social behavior. For example, they predict unrealistically low levels of voter turnout and charitable donation [37,38]. In addition, countless laboratory experiments have shown that people often behave unselfishly (e.g., sharing resources, punishing malefactors) [39•]. A number of different theoretical models attempt to capture this other-regarding behavior by modifying the standard selfish utility function to include concerns for social factors such as inequality [40,41], social welfare [42], fairness and reciprocity [43••,44•,45], or social image [46]. These models of social preferences enable pairs or groups of individuals to obtain outcomes that purely self-interested individuals cannot [47•].

In this paper, we explore models of other-regarding preferences that directly incorporate emotional rewards into the payoffs players receive in the game. These models allow players' payoffs and psychological states to depend upon their beliefs, using tools from psychological game theory [31•,32••]. As noted by Geanakoplos *et al.* [31•], and consistent with the appraisal theory approach to emotions, 'A player's emotional reactions cannot in general be independent of his expectations and of his interpretation of what he learns in a play of a game'. The psychological games approach thus requires the modeler to make precise assumptions about the appraisal triggers of emotions and the resulting consequences of those emotions for behavior. In addition, psychological game theory is well suited for modeling the theory-of-mind reasoning that is often associated with social emotions [48,49]. Both the appraisals and the action tendencies associated with guilt and anger, for example, can be modeled by adding a psychological payoff term to the standard material payoff. This approach highlights that agents face tradeoffs between psychological and material payoffs, so that emotions need not always result in a pre-programmed action.

While we focus on the behavioral predictions of the models, we believe that 'psychological payoffs' are real and can be validated by their neural and physiological correlates. We therefore also report on the results of fMRI and other studies that seek to identify physiological data that corresponds to certain emotions. Ultimately, the models we describe will either be falsified or supported via a combination of behavioral and physiological data.

Guilt

Battigalli and Dufwenberg [50••] develop a model whereby a player feels guilty to the extent his actions cause a coplayer to receive less than he expected (see also [51•,52,53]). Player A's strategy is denoted by s_A and his material payoff by π_A. A given history of the game is denoted by h Player A's guilt toward player B is determined by the function $G_{AB} = \max(\mathrm{E}_B[\pi_B|h_0] - \pi_B, 0)$, where $\mathrm{E}_B[\pi_B|h_0]$ represents B's expected payoff, calculated at the initial history (the start) of the game with respect to B's first-order beliefs and his strategy.[5] However, player A does not know what payoff player B initially expected. So player A's expected utility $\mathrm{E}_A^2[U_A]$ is a combination of material and psychological payoffs and is calculated with respect to his second-order beliefs: $\mathrm{E}_A^2[U_A(s_A)|h] = \mathrm{E}_A^2[\pi_A(s_A) - \theta_A G_{AB}(s_A)|h]$, where θ_A is a parameter reflecting Player A's sensitivity to guilt. Battigalli and Dufwenberg [50••] refer to this model as 'simple guilt'.[6]

[4] All emotions can be considered "social" to some degree as they involve communicating internal states [3,22,23].

[5] For ease of exposition, we often suppress the notation showing the dependence of each players' payoffs on strategies. More formally each players' material payoff is a function of the strategies of all coplayers, so that e.g. $\pi_A = \pi_A(s_A, s_B)$ in two player games.

[6] They also develop 'guilt from blame', where A feels guilty for letting down B only when A believes that B believes that A caused B to get less than he expected. We refer the interested reader to Battigalli and Dufwenberg [50••] for the formal model.