Guilt in Games

By Pierpaolo Battigalli and Martin Dufwenberg[1]

(January 2, 2007)

"A clear conscience is a good pillow." Why does this old proverb contain an insight? The emotion of *guilt* holds a key. Psychologists report that "the prototypical cause of guilt would be the infliction of harm, loss, or distress on a relationship partner" (Roy Baumeister, Arlene Stillwell, and Todd Heatherton 1994, p. 245; cf. June Price Tangney 1995). Moreover, guilt is unpleasant and may affect behavior to render the associated pangs counterfactual. Baumeister *et al*: "if people feel guilt for hurting their partners ... and for failing to live up to their expectations, they will alter their behavior (to avoid guilt) in ways that seem likely to maintain and strengthen the relationship". Avoided guilt is the down of the sound sleeper's bolster.

How can guilt be modeled? How is human interaction and economic outcomes influenced? We offer a formal approach for providing answers: Start with an extensive game form which associates with each end node a monetary outcome. Say that player i lets player j down if as a result of i's choice of strategy j gets a lower monetary payoff than j expected to get before play started. Player i's guilt may depend on how much i believes he lets j down. Player i's guilt may also depend on how much i believes j believes i believes he lets j down. We develop techniques for analyzing equilibria when players are motivated in part by a desire to avoid guilt.

The modeling relates a player's motivation to beliefs about beliefs about what strategic choices are made. This accords well with Baumeister et al's remark *op. cit.* about how harm and guilt depends on a failure "to live up to [others'] expectations." However, the approach does not accord well with traditional game theory, where utilities depend merely on actions chosen

[1] Pierpaolo Battigalli (corresponding author), IEP and IGIER, Bocconi University, Via Sarfatti 25, Milano 20136, Italy (email: pierpaolo.battigalli@unibocconi.it); Martin Dufwenberg, Department of Economics and Institute for Behavioral Economics, University of Arizona, Tucson 85721-0108, Arizona, USA (email: martind@eller.arizona.edu). The authors gratefully acknowledge financial support from MIUR (Battigalli) and NSF (Dufwenberg).

and not on beliefs about beliefs. The intellectual home for our exercise is rather what has been called *psychological game theory.* This framework — originally developed by John Geanakoplos, David Pearce and Ennio Stacchetti (1989) and recently extended by Battigalli and Dufwenberg (2005) (henceforth B&D) — allows players' payoffs to depend on beliefs (about choices, states of nature, or others' beliefs), as is typical of many emotions.[2]

Some previous work has considered sentiments related to those we describe for the specific context of "trust games."[3] Gary Charness and Dufwenberg (2006) coin the term "guilt aversion," which we adopt as we develop our theory for general games.

I. Game-Theoretic Preliminaries

We consider finite extensive game forms specifying monetary payoffs for each player at each end node. These payoffs describe the material concequences of players actions, not their preferences. The players' utilities will be introduced later on, in Section II.

Extensive game forms. Let N be the player set, T the set of nodes in the game tree with distinguished root t^0, and Z the set of end nodes (or terminal nodes). The set $X = T\backslash Z$ is partitioned into subsets X_i of decision nodes for each $i \in N$ and the set of chance nodes X_c. We let $\sigma_c(\cdot|x)$ denote the strictly positive chance probabilities of the immediate followers of node $x \in X_c$. In our theory it is important to represent players' information also at nodes where they are *not* active. Thus, we let the information structure of i be a partition H_i the whole set T that contains as a subcollection the standard information partition of X_i. A typical information set is denoted h. The information set containing node t is denoted $H_i(t)$. The (extended) information structure H_i satisfies *perfect recall.* We also assume that H_i is a refinement of $\{\{t^0\}, X\backslash\{t^0\}, Z\}$; this means that the players know when they are at the root of the game tree and they know when the game is over. The material consequences of players' actions are determined by functions $\mathbf{m}_i : Z \to \mathbb{R}$, $i \in N$. A typical material payoff is denoted by

[2] Jon Elster (1998) argues that a key characteristic of emotions are that they "are triggered by beliefs" (p. 49).

[3] See Peter Huang & Ho-Mou Wu (1994), Dufwenberg (1995, 2002), Dufwenberg & Uri Gneezy (2000), Michael Bacharach, Gerardo Guerra & Daniel Zizzo (2002), Guerra & Zizzo (2004), and Gary Charness and Dufwenberg (2006), Giuseppe Attanasi and Rosemarie Nagel (2006).

m_i, as in $m_i = \mathbf{m}_i(z)$. We assume that a player observes his material payoff: $\mathbf{m}_i(z') \neq \mathbf{m}_i(z'')$ implies $H_i(z') \neq H_i(z'')$. Whenever we do not explicitly specify players' terminal information, *the default assumption is that they have the coarsest terminal information consistent with perfect recall and with their material payoff.*

Pure strategies. A pure strategy s_i specifies a contingent choice for each $h \in H_i$ where i is active $(h \subset X_i)$. We find it convenient to refer to 'pure strategies' also of chance, i.e. functions $s_c : X_c \to T$ that select an immediate successor of each chance node (such strategies are chosen at random according to the mixed representation of σ_c). The set of pure strategies of i is S_i and we let $S = S_c \times \prod_{i \notin N} S_j$, $S_{-i} = S_c \times \prod_{j \neq i} S_j$. For any h and i, we let $S_i(h)$ denote the set of i's strategies allowing h, and let $S_{-i}(h) \subset S_{-i}$ denote the set of profiles s_{-i} allowing h. We use similar notation for strategies and strategy profiles consistent with a given node. A strategy profile $s \in S$ (which includes chance's strategy s_c) yields a particular end node denoted $\mathbf{z}(s)$.

Behavior strategies. We assume that players do not actually randomize, but randomized choice – in the form of behavior strategies – enter the analysis as an expression of players' beliefs. A behavior strategy for i is an array σ_i of probability measures $\sigma_i(\cdot|h)$, $h \in H_i$, $h \subset X_i$, where $\sigma_i(a|h)$ is the probability of choosing action a at informaton set h. Given σ_i we can compute the probability of each pure strategy s_i, denoted $\Pr_{\sigma_i}(s_i)$ (see Harold Kuhn, 1953). By perfect recall, one can compute the conditional probabilities $\Pr_{\sigma_i}(s_i|h)$, $h \in H_i$, even if $\Pr_{\sigma_i}(S_i(h)) = 0$.

Conditional beliefs. Conditional on each information set $h \in H_i$ player i holds an updated, or revised**,** belief $\alpha_i(\cdot|S_{-i}(h)) \in \Delta(S_{-i}(h))$ about the strategies of the co-players and of chance; we abbreviate the notation to $\alpha_i(\cdot|h)$ whenever convenient. $\alpha_i = (\alpha_i(\cdot|h))_{h \in H_i}$ is the system of *first-order* beliefs of i (note that we are including in α_i also i's beliefs about chance moves, later on we impose the restriction that they are determined by the objective probabilities σ_c). Player i also holds a *second-order* belief $\beta_i(h)$ about the first-order belief system α_j of each co-player j, a third-order belief $\gamma_i(h)$ about the second-order beliefs, and so on. For the purposes of this paper, we may assume that higher-order belief are degenerate point beliefs. Thus, with a slight abuse of notation we identify $\beta_i(h)$ with a particular array of conditional first-order beliefs $\alpha_{-i} = (\alpha_j(\cdot|h'))_{j \neq i, h' \in H_j}$. A similar notational convention applies to other higher-order beliefs. Clearly, the beliefs that i would hold at different information sets are not mutually independent. They must satisfy the *chain rule* of conditional probabilities [$S_{-i}(h'') \subset S_{-i}(h')$ and $\alpha_i(S_{-i}(h'')|S_{-i}(h')) > 0$ imply that, for all $s_{-i} \in S_{-i}(h'')$, $\alpha_i(s_{-i}|S_{-i}(h'')) = \alpha_i(s_{-i}|S_{-i}(h'))/\alpha_i(S_{-i}(h'')|S_{-i}(h'))$ and that i has the same (point) higher-order beliefs at h' and h''] and they also satisfy common certainty that such rule holds (for details, see B&D). In our equilibrium analysis we consider beliefs at most of the fourth order. Players *initial* beliefs are those held at the information set $h^0 = \{t^0\}$.

II. Two Concepts of Guilt Aversion

Given his plan of action s_j and initial first-order beliefs $\alpha_j(\cdot|h^0)$ player j forms an expectation about his material payoff: $\mathrm{E}_{s_j,\alpha_j}[m_j|h^0] = \sum_{s_{-i}} \alpha_j(s_{-j}|h^0)\mathbf{m}_j(\mathbf{z}(s_j, s_{-j}))$. For any end node z consistent with s_j, the expression $D_j(z, s_j, \alpha_j) = \max\{0, \mathrm{E}_{s_j,\alpha_j}[m_j|h^0] - \mathbf{m}_j(z)\}$ measures how much j is "let down". If at the end of the game i knew the terminal node z, the strategy profile $s_{-i} \in S_{-i}(z)$, and j's initial beliefs α_j, then he could derive how much of $D_j(z, s_j, \alpha_j)$ is due to his behavior: $G_{ij}(z, s_{-i}, \alpha_j) = D_j(z, s_j, \alpha_j) - \min_{s_i} D_j(\mathbf{z}(s_i, s_{-i}), s_j, \alpha_j)$.

Our first guilt concept draws directly on $G_{ij}(z, s_{-i}, \alpha_j)$. We say that i is affected by *simple guilt* toward j if he has belief-dependent preferences represented by a payoff function of the form

$$(1) \qquad u_i^{SG}(z, s_{-i}, \alpha_{-i}) = \mathbf{m}_i(z) - \sum_{j \neq i} \theta_{ij} G_{ij}(z, s_{-i}, \alpha_j),\ s_{-i} \in S_{-i}(z),\ \theta_{ij} \geq 0.$$

Since i does not know s_{-i} or α_{-i} and may not even observe z, u_i^{SG} does not represent a utility "experienced" by i. What we assume is that, given his first- and second-order beliefs, i tries to make the expected value of u_i as large as possible.[4] We let $(\theta_{ij})_{i,j \in N, j \neq i}$ denote the profile of players' guilt sensitivity parameters.

Whereas with simple guilt a player cares about the extent to which he lets another player down, our second formulation assumes that a player cares about others' inferences regarding

[4] (1) yields the same sequential best response correspondence as the slightly simpler function $v_i(z, s_{-i}, \alpha_{-i}) = \mathbf{m}_i(z) - \sum_{i \neq i} \theta_{ij} D_j(z, s_j, \alpha_j)$. We use (1) for two reasons: it is conceptually more appropriate (i cannot be "guilty" for behavior due to others), and expression $G_{ij}(z, s_{-i}, \alpha_{-i})$ is needed below to define our second concept of guilt.

the extent which he is willing to let them down. We model this as follows: Given his strategy s_i and initial first- and second-order beliefs $\alpha_i(\cdot|h^0)$ and $\beta_i(h^0)$, we first compute how much i expects to let j down:

$$G_{ij}^0(s_i, \alpha_i, \beta_i) = \mathrm{E}_{s_i,\alpha_i,\beta_i}[G_{ij}|h^0] = \sum_{s_{-i}} \alpha_i(s_{-i}|h^0) G_{ij}(\mathbf{z}(s_i, s_{-i}), s_{-i}, \beta_{ij}^0(h^0)) \tag{2}$$

where $\beta_{ij}^0(h^0)$ denotes the initial (point) belief of i about the initial first-order belief $\alpha_j(\cdot|h^0)$. Now suppose end node z is reached; the conditional expectation $\mathrm{E}_{\alpha_j,\beta_j,\gamma_j}[G_{ij}^0|H_j(z)]$ measures j's inference regarding how much i thinks he lets j down, or how much j "blames" i. We say that i is affected by *guilt from blame* if he dislikes being blamed. Thus i's preferences are represented by

$$u_i^{GB}(z, \alpha_{-i}, \beta_{-i}, \gamma_{-i}) = \mathbf{m}_i(z) - \sum_{j \neq i} \theta_{ij} \mathrm{E}_{\alpha_j,\beta_j,\gamma_j}[G_{ij}^0|H_j(z)],\ \theta_{ij} \geq 0 \tag{3}$$

Player i tries to make the expectation of u_i^{GB} as large as possible, given his beliefs (up to the fourth order).

When we append the functions $(u_i^{SG})_{i\in N}$ (respectively $(u_i^{GB})_{i\in N}$) to the given extensive game form we obtain a *psychological game with simple guilt* (respectively *with guilt from blame*).[5] We assume that the psychological game has complete information; in particular there is common knowledge of the psychological payoff functions (this is clearly farfetched, but incomplete information could be captured by making chance choose the parameters θ_{ij}).

III. Equilibrium Analysis

We adapt to the present framework the sequential equilibrium concept of David Kreps and Robert Wilson (1982). An *assessment* is a profile $(\sigma, \alpha, \beta, ...) = (\sigma_i, \alpha_i, \beta_i, ...)_{i\in N}$ specifying behavior strategies, first- and higher-order beliefs. Assessment $(\sigma, \alpha, \beta, ...)$ is *consistent* if there is a strictly positive sequence $\sigma^k \to \sigma$ such that for all $i \in N$, $h \in H_i$, $s_{-i} \in S_{-i}(h)$,

$$\alpha_i(s_{-i}|h) = \lim_{k\to\infty} \frac{\Pr_{\sigma_c}(s_c) \prod_{j\neq i} \Pr_{\sigma_j^k}(s_j)}{\sum_{s'_{-i}\in S_{-i}(h)} \Pr_{\sigma_c}(s'_c) \prod_{j\neq i} \Pr_{\sigma_j^k}(s'_j)}, \tag{4}$$

[5] We build on B&D's framework, not that of Geanakoplos *et al* which would not allow i's utility to depend on other players' beliefs (in contrast to (1) and (3)) or on updated beliefs (in contrast to (3)).

and higher-order beliefs at each information set are correct: for all $i \in N$, $h \in H_i$, $\beta_i(h) = \alpha_{-i}$, $\gamma_i(h) = \beta_{-i}$, $\delta_i(h) = \gamma_{-i}$, and so on.

Fix a profile of payoff functions of the form $u_i(z, s_{-i}, \alpha, ...)$ (this covers u_i^{SG} and u_i^{GB} as special cases). A consistent assessment $(\sigma, \alpha, \beta, ...)$ is a *sequential equilibrium (SE)* if each measure $\Pr_{\sigma_i}(\cdot|\cdot)$ assigns positive conditional probability only to conditional expected payoff maximizing strategies: for all $i \in N$, $h \in H_i$, $s_i \in S_i(h)$, $\Pr_{\sigma_i}(s_i|h) > 0 \Rightarrow s_i \in \arg\max_{s_i' \in S_i(h)} \mathrm{E}_{s_i', \alpha_i, \beta_i, ...}[u_i|h]$ (this sequential rationality condition is redundant, but well posed, at information sets where i is not active). If the payoff functions depend only on the end node, our definition of SE is equivalent to that of Kreps and Wilson. Adapting an existence proof from B&D, one can show that every psychological game with simple guilt, or guilt from blame, has a SE.[6]

We now list some results and examples about the relationships between SE with simple guilt and guilt from blame, as well as SE of the "material-payoff game" with payoff functions $u_i \equiv \mathbf{m}_i$. First note that in any two-player game form without chance moves, for every pure-strategy, consistent assessment $(s, \alpha, \beta, ...)$, every i and s_i',

$$G_{ij}^0(s_i', \alpha_i, \beta_i) = \max\{0, \mathbf{m}_j(\mathbf{z}(s)) - \mathbf{m}_j(\mathbf{z}(s_i', s_{-i}))\} = \mathrm{E}_{\alpha_j, \beta_j, \gamma_j}[G_{ij}^0 | H_j(\mathbf{z}(s_i', s_{-i})]. \quad (5)$$

The first equality is an immediate consequence of consistency, the second follows from consistency, perfect recall and observation of own material payoff. This implies:

Observation 1. *In any two-person, simultaneous-move game form without chance moves, for any given parameter profile $(\theta_{ij})_{i,j \in N, j \neq i}$ the pure stategy SE assessments of the psychological games with simple guilt and guilt from blame coincide.*

In other games, a SE with simple guilt need not be a SE with guilt from blame, and vice versa. To see this, consider first the following three-player simultaneous-move game form:

Example 1. Cleo (a dummy player) has 2 dollars. Ann and Bob simultaneously decide whether to *steal* from Cleo or to *abstain*. If at least one of them *steals*, Cleo is left with 0. If only one player *steals* s/he gets two dollars; if two players *steal* they get one dollar each. Ann and Bob are symmetrically affected by guilt towards Cleo: $\theta_{AC} = \theta_{BC} = \theta > 0$. If $1 < \theta < 2$,

[6] B&D argue that other solution concepts and forward induction reasoning should be explored. For space reasons we do not pursue this here.

then the strategy profile (*abstain*, *abstain*) is an SE with simple guilt but not with guilt from blame. Note the intuition: if Ann or Bob deviates from profile (*abstain*, *abstain*) and *steals*, then since Cleo observes only her material payoff of 0 she cannot be sure whom to blame. With guilt from blame, this shelters the deviator from some pangs under which a player affected by simple guilt must suffer. More formally, let $\hat{\alpha}^i = \alpha_C(a_i = steal|m_C = 0)$ be the ex post marginal probability that i deviated, as assessed by Cleo. By consistency, Cleo thinks that two deviations are infinitely less likely than one, hence $\hat{\alpha}^A + \hat{\alpha}^B = 1$ and $\hat{\alpha}^i \leq \frac{1}{2}$ for at least one i. This player has no incentive to *steal* only if $2 - \theta \times 2\hat{\alpha}^i \leq 0$, that is $\theta \geq 1/\hat{\alpha}^i \geq 2$. (Note how, with guilt from blame, off-equilibrium-path updated beliefs matter even in simultaneous game forms.)

Next consider a two-player-plus-chance game form with asymmetric information:

Example 2. Ann first observes a chance move with equally likely outcomes b or g, and then chooses *in* or *out*. If she chooses *out*, Bob (a dummy player) gets 2 dollars. If she chooses *in*, Bob's material payoff depends on chance: 0 if b, 8 if g. Ann always gets 0 dollars but is affected by guilt towards Bob. Look at the strategy profile (=strategy of Ann's) (*in*, *in*) (meaning *in* if b, *in* if g). Clearly this is not a SE with simple guilt: Bob initially expects to get $1/2 \times 0 + 1/2 \times 8 = 4$ – he is thus let down in the (expected) amount $1/2 \times 4 + 1/2 \times 0 = 2$. By deviating to (*out*, *in*) Ann can change how much Bob is let down to $1/2 \times 2 + 1/2 \times 0 = 1$. This is the unavoidable expected extent to which Bob will be let down. Thus the expected guilt associated with (*in*, *in*) is $2 - 1 = 1$, as compared to $1 - 1 = 0$ for strategy (*out*, *in*). Since material payoff is not an issue for Ann, she wants to deviate to (*out*, *in*).

Yet (*in*, *in*) is a SE with guilt from blame. It is supported by Bob's out-of-path beliefs such that if he got a material payoff of 2 then he thinks it is because Ann plays strategy (*in*, *out*).[7] The expected guilt associated to this strategy is $1/2 \times 4 + 1/2 \times 2 = 3$, and this is how much Bob blames Ann if he observes a payoff of 2 dollars. If Ann does not deviate, Bob gets a positive probability payoff (0 or 8), infers that she is indeed playing (*in*, *in*) and therefore his blame on Ann is 1, the expected guilt associated to (*in*, *in*). Therefore any deviation from (*in*, *in*) increases Bob's blame in expectation.

[7] Such a belief is consistent: consider the sequence $\sigma_A^k(in|g) = 1 - k^{-1}$, $\sigma_A^k(in|b) = 1 - k^{-2}$ for $k = 1, 2, ...$

Observation 2. *In any game form with simultaneous moves, for any parameter profile* $(\theta_{ij})_{i,j\in N, j\neq i}$*, all the pure strategy SE assessments of the material payoff game are also SE of the psychological games with simple guilt and guilt from blame.*

Proof. Fix a simultaneous game form and a SE $(s, \alpha, \beta, ...)$. Then, if i deviates from s_i he (weakly) decreases his material payoff. Given α, each player j expects to get exactly $\mathbf{m}_j(s)$; hence, if no deviations occur no player j is let down. By consistency, this implies that given (α, β) each player i (weakly) increases in expectation the absolute value of each negative component of his psychological payoff function if he deviates. Therefore a deviation by any player (weakly) decreases his total payoff. *Q.E.D.*

The following parametrized example shows that Observation 2 does not extend to sequential game forms. The example relates also to Observation 1.

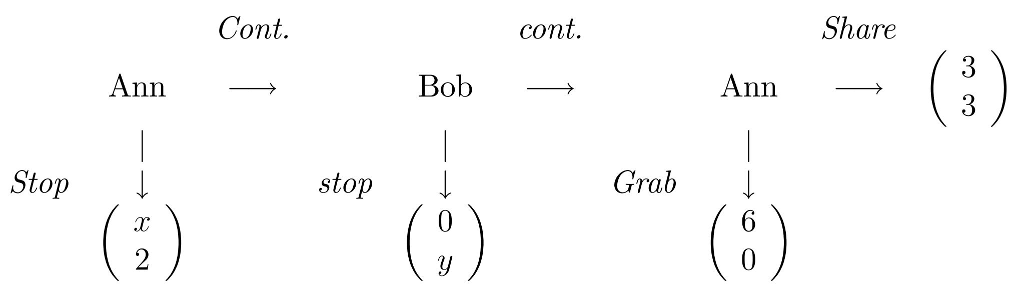

FIGURE 1. A PERFECT INFORMATION GAME FORM: $x > 0$, $y \leq 2$

Example 3. Suppose that $x > 0$, $0 < y \leq 2$. Then [(*Stop*, *Grab*), *stop*] is the only SE of the material payoff game depicted in Figure 1 and it yields outcome $(x, 2)$. This outcome is not supportable by any SE of the psychological game with simple guilt if θ_{AB} is high enough. The reason is that if Ann correctly guesses that Bob initially expects 2 dollars, then at history/node (*Cont.*, *cont.*) she would be sure to let Bob down in the amount of 2 by choosing to *Grab*. If $\theta_{AB} > \frac{3}{2}$ Ann would then prefer to *Share*. Anticipating this, Bob would continue after *Cont.*, and Ann would deviate to *Cont.* at the beginning of the game. Thus, Observation 2 does not extend to sequential game forms for simple guilt, even if we only look at equilibrium outcomes.

On the other hand, for the same parameter values [(*Stop*, *Grab*), *stop*] is a SE of the game with guilt from blame. The reason is that if Ann does not *Stop*, the blame by Bob on Ann is $\mathbf{m}_2(\mathit{Stop}) - \mathbf{m}_2(\mathit{Cont.}, \mathit{cont.}) = (2 - y)$, independently of what happens afterward, because this is how much Ann intended to let Bob down. Therefore Ann would have no incentive to *Share* if given the opportunity. This shows that Observation 1 does not extend to sequential game forms, even if we only look at equilibrium outcomes.

Now suppose that $x > 6$ and $y < 0$. The only SE of the material payoff game is [(*Stop*, *Grab*), *cont.*]. If $\theta_{AB} > \frac{3}{2}$ this is not an SE of the game with guilt from blame: since the equilibrium strategy of Bob is to choose *cont.*, in this case Ann's action in the subgame affects the guilt blamed by Bob on Ann, who would rather *Share*.

We close this section by considering an application concerning the provision of public goods. This exercise highlights the differences between the two types of guilt as well as the huge multiplicity of equilibria that may obtain with such belief-dependent motivations.

Public good games with linear technology. Each player can can contribute an integer number of dollars from zero to K. Each contributed dollar yields an increase of B dollars in the material welfare of every player, with $1/n < B < 1$. Assume a common guilt aversion parameter: $\theta_{ij} = \theta$. In the psychological game with simple guilt, if $\theta B(n-1) \geq (1-B)$, i.e. if the total guilt from withholding one dollar from the agreed upon contribution is larger than the marginal increase in material payoff, then *every* pure strategy profile is a SE.

Strategy profiles different from $(0, ..., 0)$ are harder to support as SE of the game with guilt from blame. Let $k > 0$ be the number of players, called "donors", that give a positive contribution. According to our default assumption each player only observes his material payoff and action, which creates an ex post inference problem: a shortfall of dB (dollar) units from the expected amount of public good can be "blamed" on any one of the players that are supposed to contribute at least d. A given strategy profile has the best chance to be supported as an equilibrium if non-deviators observing a shortfall of B assess the same probability that any of the k donors (or $k-1$ *other* donors) gave one dollar less. In this case the total guilt blamed on the actual unilateral deviator by other donors for a shortfall of B dollars is $(k-1)B/(k-1) = B$ (or 0 if $k = 1$, i.e. if there are no other donors) and the total guilt blamed by non-donors is $(n-k)B/k$. Thus a donor does not want to withhold one dollar if and only if $\theta B[I_{k>1}(k) + (n-k)/k] \geq (1-B)$, where $I_{k>1}(k) = 0$ if $k = 1$ and $I_{k>1}(k) = 1$ otherwise. This is a necessary condition for a SE with k donors, and it is also sufficient if each donor gives the same amount. We obtain the same condition as with simple guilt if $k = 1$ because in this case the deviating donor is identified and "fully blamed". Thus, profiles with more asymmetric contributions are easier to support because they mitigate the inference problem.

Now consider the following variation of the public good game form. Each additional dollar given for the public good yields B additional (dollar) units of public good with probability $p < 1$ and no social benefit with probability $1-p$, independently of other contributions (assume $1/n < pB < 1$). If $\theta p^2 B(n-1) \geq (1-B)$ every pure strategy profile is a SE with simple guilt. But only $(0, ..., 0)$ is a pure SE of the game with guilt from blame. The reason is that every shortfall from the expected amount of public good is blamed on bad luck, thus there is no incentive to give any agreed upon positive contribution (however, for large enough θ there are mixed SE with positive expected contributions).

IV. Concluding Remarks

We develop a general theory of guilt aversion and show how to solve for sequential equilibria. We hope the approach will prove useful for a variety of applications concerning economic situations where it seems plausible that decision makers are affected by guilt.

To end on a more general note, psychological game theory provides the intellectual home for our approach. Few previous applications of that framework exist. The most prominent examples concern kindness-based reciprocity (*e.g.* Matthew Rabin 1993, Dufwenberg and Georg Kirchsteiger 2004), anxiety (Andrew Caplin and John Leahy 2004; cf. Caplin and Leahy 2001), and social respect (Douglas Bernheim 1994, Dufwenberg and Michael Lundholm; these authors do not explicitly refer to psychological games but their work fits the framework of B&D). The usefulness of psychological game theory for studying these diverse kinds of motivation augurs well for the framework's potential for analyzing also other phenomena including disappointment, regret, anger, surprise, shame, and joy.

References

[1] **Attanasi, Giuseppe and Nagel, Rosemarie.** "Actions, Beliefs and Feelings: An Experimental Study on Dynamic Psychological Games", mimeo, Bocconi University, Universitat Pompeu Fabra, 2006.

[2] **Bacharach, Michael; Guerra, Gerardo and Zizzo, Daniel J.** "Is Trust Self-Fulfilling? An Experimental Study," mimeo, 2001.

[3] **Battigalli, Pierpaolo and Dufwenberg, Martin.** "Dynamic Psychological Games", IGIER working paper 287, Bocconi University, 2005.

[4] **Baumeister, Roy F.; Stillwell, Arlene M. and Heatherton, Todd F.** "Guilt: An interpersonal approach," *Psychological Bulletin*, 1994, *115*(2), pp. 243-267.

[5] **Bernheim, B. Douglas.** "A Theory of Conformity," *Journal of Political Economy,* 1994, *102*(5), pp. 841-77.

[6] **Caplin, Andrew S. and Leahy, John V.** "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics,* 2001, *116*(1), 55-79.

[7] **Caplin, Andrew S. and Leahy, John V**. "The Supply of Information by a Concerned Expert," *Economic Journal,* 2004, *114*(497, July), pp. 487-505.

[8] **Charness, Gary and Dufwenberg, Martin.** "Promises and Partnership", *Econometrica*, 2006, *74*(6), pp. 1579-1601.

[9] **Dufwenberg, Martin.** "Time-Consistent Wedlock with Endogenous Trust", in Doctoral Dissertation, Uppsala University, 1995.

[10] **Dufwenberg, Martin**. "Marital Investment, Time Consistency and Emotions," *Journal of Economic Behavior and Organization, 48*(1), pp. 57-69.

[11] **Dufwenberg, Martin and Gneezy Uri**. "Measuring beliefs in an experimental lost wallet game," *Games and Economic Behavior,* 2000, *30*(2), pp. 163-182.

[12] **Dufwenberg, Martin and Kirchsteiger Georg**. "A Theory of Sequential Reciprocity," *Games and Economic Behavior,* 2004, *47*(2), 268-298.

[13] **Dufwenberg, Martin and Lundholm Michael.** "Social Norms and Moral Hazard," *Economic Journal,* 2001, *111*(473, July), pp. 506-525.

[14] **Elster, Jon.** "Emotions and Economic Theory," *Journal of Economic Literature,* 1998, *36*(1), pp 47-74.

[15] **Geanakoplos, John; Pearce, David G. and Stacchetti Ennio**. "Psychological Games and Sequential Rationality," *Games and Economic Behavior,* 1989, *1*(1), 60-79.

[16] **Guerra, Gerardo and Zizzo, Daniel J.** "Trust Responsiveness and Beliefs," *Journal of Economic Behavior and Organization,* 2004, *55*(1), pp. 25-30.

[17] **Huang, Peter H. and Wu, Ho-Mou.** "More Order without More Law: A Theory of Social Norms and Organizational Cultures," *Journal of Law, Economics and Organization,* 1994, *10*(2), pp. 390-406.

[18] **Kreps, David M. and Wilson, Robert B.**. "Sequential Equilibrium," *Econometrica,* 1982 *50*(4), pp. 863-894.

[19] **Kuhn, Harold W.** (1953): "Extensive Games and the Problem of Information," in *Contributions to the Theory of Games* II, ed. by H. W. Kuhn and A. W. Tucker. Princeton: Princeton University Press, pp. 193-216.

[20] **Rabin, Matthew**. "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 1993, *83*(5), 1281-1302.

[21] **Tangney, June Price.** "Recent Advances in the Empirical Study of Shame and Guilt," *American Behavioral Scientist*, 1995, *38*(8), pp. 1132-1145.