Strategic Guilt Induction*

Eric Cardella†

November 6, 2011

Abstract

Guilt averse agents are motivated to meet the expectations of others, even at the expense of their own material payoff. Several experimental studies have found results consistent with guilt averse motivations in games. However, there are strategic implications of guilt aversion, which can impact economic outcomes in important ways, that have yet to be explored. I introduce a game that admits the possibility for agents to induce guilt upon others in a manner consistent with the method posited by Baumeister, Stillwell, and Heatherton (1994). This game enables me to experimentally test whether agents attempt to strategically inducing guilt upon others, and whether guilt induction is an effective mechanism for influencing the behavior of others. Additionally, the design enables me to test whether agents exhibit higher degrees of trust when they are given an opportunity to induce guilt. Furthermore, I appeal to the Battigalli and Dufwenberg (2007) model of simple guilt and derive conditions under which effective guilt induction can be supported as an equilibrium of the psychological game considered.

Keywords: guilt aversion, trust, psychological game theory, experiment

JEL Codes: C72, C91, D03, D80

*I would like to thank my dissertation advisor Martin Dufwenberg for helpful comments. I would also like to thank Pierpaolo Battigalli, Anna Breman, Gary Charness, Price Fishback, Paul Heidhues, Tamar Kugler, John Wooders, and conference participants at the University of Arizona for additional comments. This project was funded by the Economic Science Laboratory and the National Science Foundation (grant # SES-0819780).

†Department of Economics, University of Arizona, McClelland Hall 401, PO Box 210108, Tucson, AZ 85721-0108, Telephone: (858) 395-6699; Email: ecardell@email.arizona.edu.

1 Introduction

A growing body of experimental literature suggests that economic agents making strategic decisions may not be solely motivated to maximize their own material payoffs. Battigalli and Dufwenberg (2007) (B&D henceforth) develop a model of "simple guilt" where agents suffer disutility, in the form of guilt, from letting others down.[1] Their model posits that agents who are averse to feeling guilty may be motivated to comply with the expectations of other agents, even at the expense of their own material payoff. Behavior consistent with guilt aversion has been documented in several experimental studies.[2] In light of the results from these experimental studies, it is important to consider the richer set of interpersonal strategic implications that can arise if agents are motivated by guilt aversion.

The guilt aversion of one agent can influence the strategic behavior of other agents in important ways. In certain strategic settings, the possibility may arise for agents to behave opportunistically and exploit the guilt aversion of others. Charness and Dufwenberg (2006), in their concluding remarks, point to such a possibility and raise the question, "do people manipulate the guilt aversion of others in self-serving ways" (p. 1595)?[3] Given the opportunity, and incentive, agents could attempt to influence the behavior of guilt averse others by strategically inducing guilty feelings upon them. Consequently, guilt averse agents would be more motivated to respond in kind to meet the expectations of those who induce guilt upon them, thus alleviating the possible guilty feelings. These interpersonal implications of guilt aversion can impact strategic decision making and, consequently, economic outcomes in important ways that have yet to be explored.

The goal of this study is to explore these interpersonal implications of guilt aversion in a strategic economic setting. Specifically, this paper investigates the following three questions: First, do economic agents attempt to exploit the guilt aversion of other agents in self-serving ways by strategically inducing guilt upon those other

[1] Dufwenberg (2002) develops a model of guilt applied to a specific game, the "Marital Investment Game", while B&D develop a model of guilt for a general class of games. B&D also model a second form of guilt, "guilt from blame", however, in this paper I will only consider simple guilt; therefore, for the remainder of the paper when I refer to the guilt model of B&D, I am implicitly refering to the model of simple guilt. The B&D model is an application of the authors' more general framework developed in Battigalli and Dufwenberg (2009), which extends the psychological game theory framework pioneered by Geanakoplos, Pearce, and Stacchetti (1989).

[2] Dufwenberg and Gneezy (2000), Charness and Dufwenberg (2006), Bacharach, Guerra, and Zizzo (2007), and Reuben, Sapienza, and Zingales (2009) consider variations of two-player experimental "trust" games (Berg, Dickhaut, and McCabe 1995) and find experimental evidence consistent with guilt aversion. Similarly, Dufwenberg, Gächter, and Hennig-Schmidt (2011) find evidence consistent with guilt aversion in an experimental public goods game. On the contrary, Ellingsen, Johannesson, Tjotta, and Torsvik (2010) also consider various 2-player trust games, but their study reveals little experimental support for guilt aversion.

[3] Rabin (1993) raises a similar question, albeit in a more general context and not specifically in reference to guilt, about whether players may be able to "force" emotions in a sequential move game.

agents? Second, is strategic guilt induction an effective mechanism for influencing the behavior of other agents? Third, is the strategic behavior of agents affected by having the opportunity to induce guilt upon another agent?

Although previously unexplored in the economic literature, the interpersonal implications of guilt aversion have been recognized and documented in the psychology literature (Vangelisti, Daly, and Rudnick 1991; Baumeister, Stillwell, and Heatherton 1994, 1995 (BSH henceforth); Tangney and Fischer 1995). In particular, BSH (1994) argue that one of the primary functions of guilt is to motivate others to behave more pleasingly. In their study, BSH (1994) note that "we observed ample evidence of the hypothesized function of guilt as an interpersonal influence technique: People induced guilt to get another person to comply with their wishes" (p. 249). Similarly, Vangelisti et al. argue that people induce guilt "primarily to achieve their own end-to persuade their listeners to do or not to do something" (p. 33). These psychology studies provide valuable insights regarding the functions of guilt in social relationships by drawing insights from personal narratives and surveys. However, the interpersonal functions of guilt may not be restricted to social interactions; guilt may also function as an "interpersonal influence technique" in strategic economic interactions. An incentivized experimental game provides a suitable platform for investigating these interpersonal implications of guilt aversion in economic settings.

The following example illustrates an economic setting, perhaps familiar to some, where strategic guilt induction could be employed as a means of influencing the behavior of another in a self-serving way.

Example 1 *You are a homeowner and you hire a private contractor to complete an addition on your home by a pre-specified date. Shortly thereafter, the contractor has an opportunity to take-on another well-paying job. In order for the contractor to take-on and complete the new job on time, he must delay the completion of your addition by several weeks. The contractor, however, is unaware that you had planned for your in-laws to come stay in the new addition shortly after the original agreed upon completion date. Hence, the contractor is unaware of how much you would suffer if the addition is not completed on time and your in-laws are not able to visit and stay in the new addition. Anticipating the guilt aversion of the contractor, you decide to inform the contractor that your in-laws are coming into town and how much you would suffer as a result of the addition not being completed by the original agreed upon date. To avoid the guilt the contractor would feel by delaying the projects completion, the contractor decides to work nights and weekends to complete the addition on time. As a result of your strategic guilt induction, the addition is completed on time and you get to spend a few wonderful weeks with your in-laws!*

This example highlights a specific economic setting where strategic guilt induction could be used to influence the behavior of a guilt averse other. However, strategic guilt induction could be used to influence the behavior of others and impact outcomes in other, more general, economic settings. In the workplace, managers could possibly mitigate employee shirking by conveying to employees how their sub-standard effort adversely affects other employees.[4] In contracting environments, guilt induction may allow a disadvantaged party to influence the behavior of an advantaged counter-party. In particular, a contracted firm making specific investments could possibly thwart opportunist re-contracting and hold-up by conveying to the counter-party firm the loss in profits associated with such a hold-up. In academia, guilt induction could help an assistant professor receive a more timely review decision on a submitted paper. The assistant professor could possibly motivate a guilt averse journal editor into making a more prompt review decision by *gently* informing the editor, at the time of submission, that his/her tenure review is approaching and a lengthy review period could hinder his/her tenure prospects.[5] Many economic settings, like those mentioned, permit the possibility to induce guilt upon others. A deeper understanding of the interpersonal implications of guilt aversion is required to accurately predict how guilt aversion impacts outcomes in such settings.

In order to investigate whether agents strategically induce guilt upon others, and the effectiveness of strategic guilt induction, it is crucial to first identify how agents attempt to induce guilt upon others. For this, I draw valuable insights from BSH (1994) who posit the following method for how people induce guilt in others: "If Person A wants Person B to do something, A may induce guilt in B by conveying how A suffers over B's failure to act in the desired fashion" (p. 247). This method for inducing guilt implicitly requires that (i) Person A have private information about his own well-being, and (ii) have the possibility to convey such private information to Person B. Therefore, the chosen strategic setting must feature both private payoff information and the ability to convey the private payoff information. However, the previous studies that have investigated guilt aversion in strategic settings mostly consider variations of 2-player "trust" games that do not feature private payoff information or the possibility to convey the private payoff information.[6] Hence, a new game is warranted that provides a rich enough strategic structure to allow agents the opportunity to induce guilt upon others.

I employ an experimental design that uses a 2-player trust game featuring both private payoff information and the opportunity to convey the private information. In the game, the privately informed first-mover (Player A) is given an opportunity to

[4]Sub-standard effort by employees is likely to result in lower profits for a firm. Assuming that bonuses are increasing in firm profits, then lower profits would lead to lower bonuses for all employees. Thus, shirking by one employee could adversely affect the well-being of other emplyees.

[5]This example was inspired by an editor who revealed to me that some assistant professors do actually inform the editor, at the time of submission, about such an upcoming tenure review!

[6]Fong, Huang, and Offerman (2007) consider a trust game with private information. However, the authors incorporate private information as a means of testing their model of guilt driven reciprocity. Additionally, their trust game features private information for the second mover, while I consider a trust game with private information for the first mover.

convey to the second-mover (Player B) how low his/her payoff would be as a result of Player B choosing an action that is undesirable for Player A. This effectively allows Player A to convey to Person B how much Player A would "suffer" over Player B's failure to act in the desired fashion, which is consistent with the BSH (1994) method of guilt induction.[7] This game allows me to derive hypotheses to test whether Player A attempts to strategically induce guilt upon Player B, and whether Player B is susceptible to strategic guilt induction. The design also uses a second, related, trust game that does not feature an opportunity for Player A to induce guilt upon Player B. This second game provides a baseline trust measure for Player As, which then allows me to derive a hypothesis to test if Player As are more trusting when they have an opportunity to induce guilt.

Furthermore, I theoretically explore the connection between the psychological insights of BSH (1994) regarding the interpersonal implications of guilt and the B&D model of guilt. In doing so, I apply the B&D model of guilt to the proposed experimental trust game. I derive conditions on beliefs under which the method for guilt induction, as posited by BSH (1994), is consistent with B&D. That is, I derive conditions on beliefs for which the B&D model predicts that Person A can induce guilt in Person B by conveying how low his payoff would be as a result of Person B choosing an undesirable action. Subsequently, I show that effective guilt induction can be supported as a sequential equilibrium of the proposed game under the framework of B&D.

2 Hypothesis Development

In this section, I first introduce the two trust games from which the research hypotheses are developed. I refer to both games as trust games because they feature a payoff structure indicative of the broader class of trust games. Namely, a game where the first mover has an opportunity to choose an action that creates the possibility of mutual benefit, if the other person cooperates, but a risk of lower payoffs to oneself if the other person defects. Such an action taken by the first mover is referred to as a trusting action (Cox 2004). Trust games, in general, allow for the possibility of guilty feelings, which make them a suitable platform for developing hypotheses to test the research questions of this study.

[7] I use the term "guilt induction" when referring to Person A's strategic attempt to influence the behavior of Person B by conveying how low his payoff would be given a future undesirable action of Person B. I do this to remain consistent with the psychological intuition and terminology outlined by BSH (1994). However, in relation to the B&D model of guilt aversion, it is more pedagogical to think of this strategic behavior from Person A as "counterfactual" guilt induction. Essentially, Person A is trying to increase the amount of guilt that Person B *would* feel as a result of choosing an action that is undesirable for Person A. This makes the guilt counterfactual in the sense that Person B may never experience the guilt if he chooses an action that complies with Person A's desires.

2.1 Trust Game with Uncertain Payoffs – Γ_{UPT}

Γ_{UPT} is a 2-player, sequential move game. Γ_{UPT} begins with the first mover, Player A, choosing between *In* or *Out.* If Player A chooses *Out,* then the game ends; Player A receives a payoff of 6, and Player B receives a payoff of 2. If Player A chooses *In*, then Player B is called upon to move. Player B must choose between *Left* or *Right.* If Player B chooses *Right,* then the game ends; Player A receives a payoff of 10, and Player B receives a payoff of 4. If Player B chooses *Left,* then the game ends; Player A receives a payoff of X, and Player B receives a payoff of 6. X is a random variable where $prob(X = 0) = 1/2$ and the $prob(X = 6) = 1/2$. At the start of the game, the distribution of X is known to both players. The extensive form of Γ_{UPT} is depicted below in Figure 1.

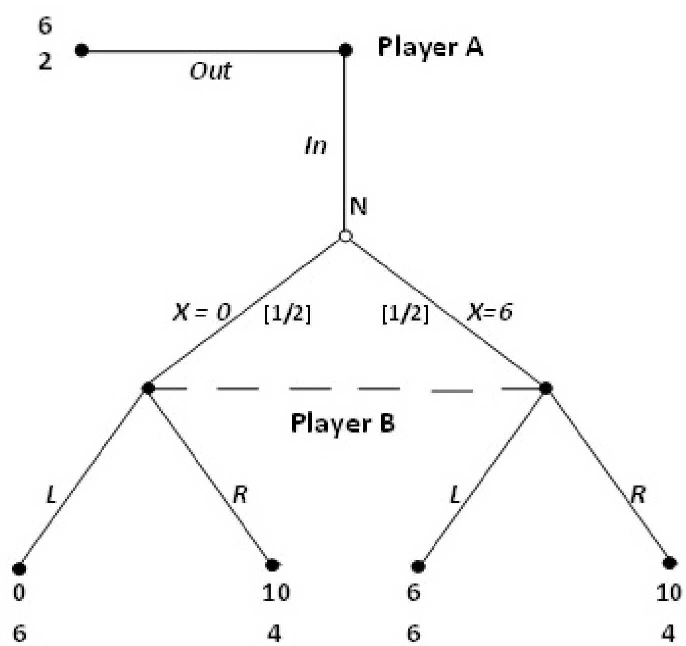

FIGURE 1: EXTENSIVE FORM OF Γ_{UPT}

2.2 Trust Game with Private Payoffs – Γ_{PPT}

Γ_{PPT} features a similar strategic structure and payoffs to those of Γ_{UPT}, with two important differences. First, Γ_{PPT} features an opportunity for Player A to become privately informed about the value of X. Second, Γ_{PPT} features an additional stage where Player A has the opportunity to credibly convey his private information about the value of X to Player B.

Γ_{PPT} begins analogously to Γ_{UPT} with Player A first choosing between *In* or *Out.* If Player A chooses *Out*, the game ends; Player A receives a payoff of 6, and Player B receiving a payoff of 2. If Player A choose *In,* Nature then decides whether Player A becomes privately informed about the value of X. With $prob = 4/5$, Nature *Reveals (Rev)* the value of X to Player A, and with $prob = 1/5$, Nature does *Not Reveal (Not Rev)* the value of X to Player A.

If the value of X is revealed to Player A, then an additional stage arises where Player A must decide whether to credibly *Convey (C)* on *Not Convey (NC)* the value of X to Player B *before* Player B gets the move. Upon getting the move, Player B must then decide between *Left* or *Right.* Analogous to Γ_{UPT}, if Player B chooses *Right,* then the game ends; Player A receives a payoff of 10, and Player B receives a payoff of 4. If Player B chooses *Left,* the game ends; Player A receives a payoff of X, and Player B receives a payoff of 6. The extensive form of Γ_{PPT} is depicted in Figure 2. To simplify the extensive form, the two moves by Nature, determining the value of X and determining whether the value of X is revealed to Player A, have been combined into one move.

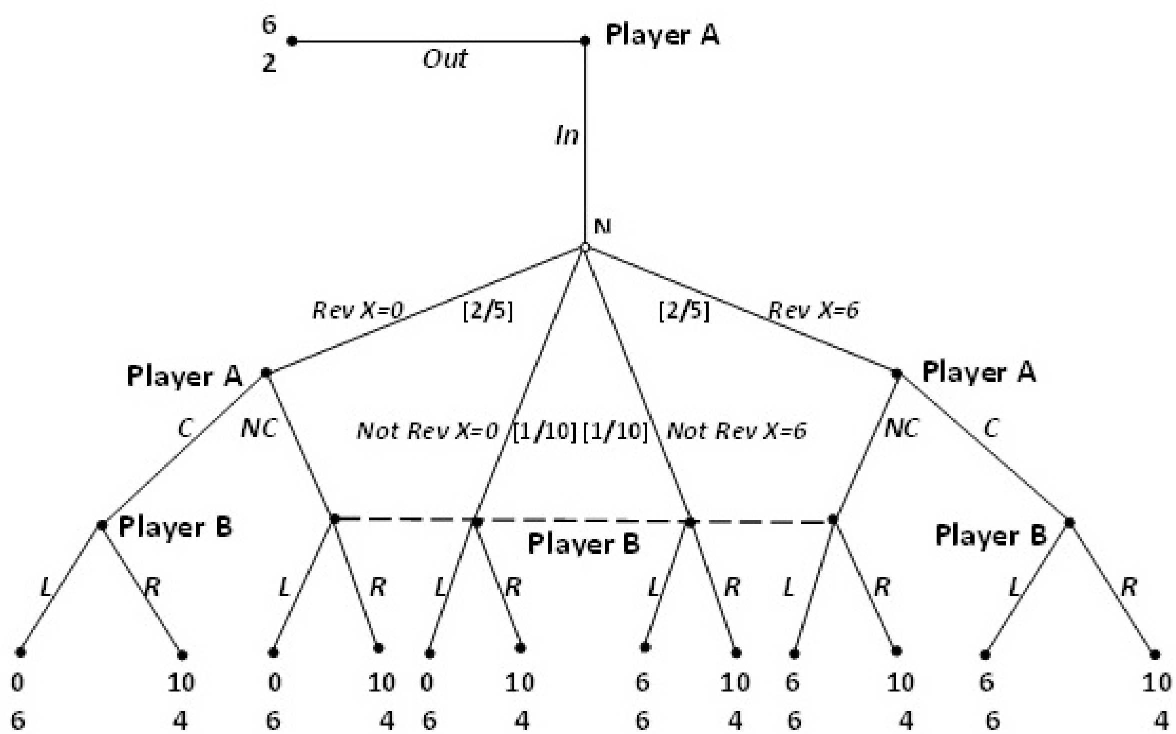

FIGURE 2: EXTENSIVE FORM OF Γ_{PPT}

If players are "selfish", i.e., act to maximize their own material payoff, then the unique equilibrium outcome in both Γ_{UPT} and Γ_{PPT} is Player A chooses *Out.* To see this, note that it is dominant for Player B to choose *Left* in both Γ_{UPT} and Γ_{PPT}. As a result, it is sequentially rational for a selfish Player A to choose *Out* in both Γ_{UPT} and Γ_{PPT}, and the unique equilibrium outcome is Player A chooses *Out.*[8] The inclusion of private information and the additional conveyance stage for Player A in Γ_{PPT} has no impact on the equilibrium outcome assuming selfish players. However, it is exactly these two features of Γ_{PPT} that will allow me to derive testable hypotheses regarding whether agents attempt to induce guilt, and its subsequent effectiveness.

Before I proceed, I first highlight two important features of Γ_{PPT} that will be relevant for the upcoming derivation of the research hypotheses and the application of the B&D model of guilt. First, Player A must choose between *In* or *Out* before possibly becoming informed about the value of X. This eliminates any possible signaling value regarding the value of X inferred from Player A's decision of *In* or *Out.* Additionally, the value of X is only revealed to a Player A who choose *In* with $prob = 4/5$. Therefore, if the value of X is not conveyed to Player B, then Player B is unable to perfectly distinguish between whether the value of X was not revealed to Player A, or the value of X was revealed and Player A chose to *Not Convey.* As a result, if the value of X is not conveyed to Player B, then Player B's expectation of the value of X will be strictly greater than zero and strictly less than six. More precisely, $\widehat{m}_A \in [1, 5]$ where $\widehat{m}_A$ denotes Player B's expectation of X, conditional on the value of X not being conveyed.[9] The notational choice of $\widehat{m}_A$ anticipates the upcoming application of the B&D model of guilt aversion.

2.3 Research Hypotheses

The first motivation of this study is to investigate whether agents attempt to strategically induce guilt upon others. Recall that BSH (1994) posit that a person can induce guilt upon another by conveying to that person how one suffers over that person's failure to act in the desired fashion. Let us consider how this method of guilt induction relates to Γ_{PPT}. Conditional on choosing *In*, a Player A would "desire" Player B

[8] In Γ_{PP}, there is multiplicity of Perfect Baysian Equilibria for selfish players that depend on the specification of Player B's beliefs at the information set where no information is conveyed regarding the value of X. However, regardless of Player B's beliefs, it is rational for him to choose *Left* and subsequently, it is sequentially rational for Player A to choose *Out.* Therefore, the unique equilibrium outcome of Γ_{PP} is the game ending with Player A choosing *Out.*

[9] To see this, note that the largest expectation that Player B could hold regarding the value of X is if he thinks that only a Player A who learned that $X = 6$ would choose to *Not Convey.* In this case, $\widehat{m}_A$ is bounded above by $\frac{1}{3} \cdot E[X] + \frac{2}{3} \cdot 6 = \frac{1}{3} \cdot 3 + \frac{2}{3} \cdot 6 = 5$. Here, $\frac{1}{3}$ and $\frac{2}{3}$ represent the updated probabilities, via Bayes' rule, that Player A did not learn the value of X and Player A learned that the value of $X = 6$, respectfully. By a similar argument, the smallest expectation that Player B could hold regarding the value of X is if he thinks only a Player A who learned $X = 0$ would choose to *Not Convey.* In this case, $\widehat{m}_A$ is bounded below by $\frac{1}{3} \cdot E[X] + \frac{2}{3} \cdot 0 = \frac{1}{3} \cdot 3 + \frac{2}{3} \cdot 0 = 1$. Therefore, $\widehat{m}_A \in [1, 5]$.

to choose *Right* yielding him a payoff of 10 compared to a payoff of $X < 10$ if Player B were to choose *Left*. The extent to which Player A would "suffer" from Player B's failure to choose *Right* depends on the value of X; the lower X the lower Player A's payoff and the more Player A would suffer. Given the opportunity, the BSH (1994) method for inducing guilt would posit that Player A could induce guilt upon Player B by "conveying" to Player B a *low* value of X, i.e., by conveying to Player B how much he would suffer if Player B were to choose *Left.*

If the value of X is revealed to Player A, after having chosen *In* in Γ_{PPT}, then Player A must decide whether to convey the value of X to Player B. In the case where $X = 0$ is revealed, if Player A chooses to *Convey* $X = 0$, then Player B will know that if he chooses *Left,* Player A will receive a payoff of $X = 0$. Whereas, if Player A chooses to *Not Convey* $X = 0$, then Player B will think that if he chooses *Left*, Player A will receive a payoff of $\widehat{m}_A \in [1, 5]$. Analogously, consider the case where the value of $X = 6$ is revealed. If Player A chooses to *Convey* $X = 6$, then Player B will know that if he choose *Left,* Player A will receive a payoff of $X = 6$. Whereas, if Player A chooses to *Not Convey* $X = 6$, Player B will think that if he chooses *Left,* Player A will receive a payoff of $\widehat{m}_A \in [1, 5]$. Hence, Player B will know that Player A suffers more from his choice of *Left* when $X = 0$ is conveyed, compared to when the value of X is not conveyed. Similarly, Player B will know that Player A suffers more from his choice of *Left* if the value of X is not conveyed, compared to if $X = 6$ is conveyed. Therefore, an opportunistic Player A who is attempting to strategically induce guilt upon Player B would *Convey* $X = 0$, and *Not Convey* $X = 6$. This leads to the first testable hypothesis:

H1: The proportion of Player As who *Convey* $X = 0$ in Γ_{PPT} is larger that the proportion of Player As who *Convey* $X = 6$.[10]

Later, I will apply the B&D model of guilt to Γ_{PPT}. I will show that, under certain conditions on the beliefs of Player B, the B&D model posits that Player B will suffer more guilt from choosing *Left* if $X = 0$ is conveyed, compared to if the value of X is not conveyed. Similarly, Player B will suffer more guilt from choosing *Left* if the value of X is not conveyed, compared to if $X = 6$ is conveyed. This implies that the hypothesis that Player A can attempt to induce guilt upon Player B by choosing to *Convey* $X = 0$ and *Not Convey* $X = 6$ in Γ_{PPT} is consistent with the B&D model of guilt. I will further show that inducing guilt by choosing to *Convey* $X = 0$ and *Not Convey* $X = 6$ can be supported in an equilibrium of Γ_{PPT} if Player B is sufficiently sensitive to feeling guilty.

[10]BSH (1994) note that guilt results from hurting others, thus a person may suffer guilt from inducing guilt. The authors refer to this type of guilt from inducing guilt as "metaguilt". It is possible that Player A could therefore suffer metaguilt from conveying $X = 0$ and not conveying $X = 6$. These metaguilty feelings could then motivate Player A to *Not Convey* $X = 0$ and to *Convey* $X = 6$. However, this potential metaguilt effect is in the opposite direction of H1, which would not give rise to a possible confounding effect.

The second motivation of this study is to investigate whether guilt induction is effective. That is, are agents more motivated to respond in the desired fashion after guilt has been induced upon them? BSH (1995) posit that after Person A has induced guilt in Person B, "Person B finds the guilt aversive and, to escape from guilt, complies with A's wishes" (pp. 247). It is also possible, however, that a guilt averse Player B will recognize that Player A is trying to manipulate his behavior by "guilting" him, which can result in Player B being more motivated to choose the unkind action of *Left* in response to Player A's attempted guilt induction. BSH (1994) (1995) document this potential "cost" of guilt induction by arguing that target of guilt induction (Person B) might feel resentment and be motivated to respond negatively toward the guilt inducer (Person A).[11] Hence, attempted guilt induction by Player A may be counterproductive as a means of motivating Player B to choose *Right* as it may foster more selfish behavior and motivate Player B to choose *Left*, contrary to Player A's intended motivation.

In relation to Γ_{PPT}, Player B complying with Player A's wishes corresponds to Player B choosing *Right.* Therefore, if guilt induction by Player A is an effective influence mechanism, then Player B would by more motivated to choose *Right* after Player A induces guilt by choosing to *Convey* $X = 0$ and *Not Convey* $X = 6$. This leads to the following two testable hypotheses:

H2a: The proportion of Player Bs choosing *Right* in Γ_{PPT} after $X = 0$ is conveyed is larger than when the value of X is not conveyed.

H2b: The proportion of Player Bs choosing *Right* in Γ_{PPT} after the value of X is not conveyed is larger than when $X = 6$ is conveyed.[12]

The third motivation of this study is to investigate whether having the opportunity to induce guilt fosters more trusting behavior. Comparing Γ_{UPT} and Γ_{PPT}, we can see that the differences between Γ_{UPT} and Γ_{PPT} are the possibility for Player A to become privately informed about the value of X, and the ability to convey the learned value of X to Player B prior to Player B's move. As I have argued, it is these differences that provide the opportunity to Player A to induce guilt in Player B. Therefore, if having an opportunity to induce guilt fosters more trusting behavior, then Player As would be more motivated to choose *In* in Γ_{PPT}, compared to Γ_{UPT}.[13] This leads to the following testable hypothesis:

[11] Alternatively, this intuition could also be viewed as some kind of negative reciprocity. A guilt averse Person B may view Person A's attempted guilt induction as an "unkind" action. As a result, Person B may be more motivated to reciprocate with the "unkind" action of choosing *Left* back toward Person A.

[12] Note, if all Player As attempt to induce guilt, then the value of $X = 6$ would never actually be conveyed to Player B. Thus, there would exist not data on the proportion on Player Bs choosing *Right* after $X = 6$ was conveyed. However, as we will see in the results section, some Player As do convey $X = 6$ so there exists data to test H2b as stated.

[13] The idea that *Right* constitutes a trusting action for Player A in Γ_{UPT} and Γ_{PPT} is consistent with the behavioral definitions of trust presented in Cox (2004) and Fehr (2009).

H3: The proportion of Player As who choose *In* when playing Γ_{PPT} is larger than the proportion of Player As who choose *In* when playing Γ_{UPT}.

I conclude this section by addressing a plausible, and widely known, alternative behavioral motivation for Player B – inequality aversion. The inequality aversion models (Fehr and Schmidt 1999; Bolton and Ockenfels 2000) posit that agents are averse to unequal outcomes, both advantageous and disadvantageous. In many trust games, inequality aversion can motivate the second mover (Player B) to choose the more kind action (*Right*), as it generally results in a more equal outcome. However, the experimental design anticipates and controls for possible inequality aversion of Player Bs. Specifically, an inequality averse Player B playing Γ_{PPT} always prefers to choose *Left*, regardless of the value of X and Player B's sensitivity to unequal outcomes.[14] It follows that Player A's conveyance decision regarding the value of X in Γ_{PPT} cannot be explained by any strategic considerations regarding the inequality aversion of Player B.[15]

3 Experimental Design and Procedure

I test the three main research hypotheses of this paper (H1-H3) using the following two experimental treatments:

(1) UPT Treatment Subjects played Γ_{UPT} one time as either Player A or Player B, where the payoffs from Γ_{UPT} corresponded 1:1 with the monetary payoffs in the experiment.

(2) PPT Treatment Subjects played Γ_{PPT} one time as either Player A or Player B, where the payoffs from Γ_{PPT} corresponded 1:1 with the monetary payoffs in the experiment.

The data was collected from experimental sessions that were conducted in the Economic Science Laboratory (ESL) at the University of Arizona in April 2011. All

[14] The intuition is relatively straight forward. When Player B chooses *Left,* his payoff increases *and* the outcome becomes less unequal. Therefore, it is optimal for an inequality averse Player B to always choose *Left.* For example, in the Fehr and Schmidt (1999) model, an inequality averse Player B always prefers the payoff vector $(X, 6)$ to the payoff vector $(10, 4)$ $\forall X \in [0, 6]$. To see this note that for the extreme case where $X = 0$, we have that $6 - \beta(6 - 0) > 4 - \alpha(10 - 4)$ $\forall \beta \in [0, 1)$ and $\alpha \geq \beta$, where the LHS represents an inequality averse Player B's utility from choosing *Left* and the RHS represnts the utiility from choosing *Right.* For the other extreme case where $X = 6$, we have that $6 > 4 - \alpha(10 - 4)$ $\forall \alpha \geq 0$, where the LHS represents an inequality averse Player B's utility from choosing *Left* and the RHS represnts the utiility from choosing *Right.* The assumputions that $\beta \in [0, 1)$ and $\alpha \geq \beta$ are assumed apriori in the model.

[15] It is possible that Player A may also be motivated by inequality aversion. Although I do not consider this in the derivation of the hypotheses, I do address the possible inequality aversion preferences of Player A when I present and discuss the experimental results.

sessions were computerized and the software was programmed using Z-tree.[16] The subject pool consisted of undergraduates who were recruited via an online database. A total of 14 sessions were conducted using a total of 312 subjects comprising 156 groups. Of the 156 groups, 111 were assigned to the PPT Treatment and the remaining 45 were assigned to the UPT Treatment. Each session lasted approximately 25 minutes and subjects earned an average of $9.54 USD (including a $5 show-up payment).

Subject were randomly assigned to their player role (either Player A or Player B), and then randomly and anonymously matched with a subject of the opposite player role. A copy of the experimental instruction are presented in the Appendix. Upon the completion of the game, the decisions of each player, the corresponding outcome, the profit to each player, and the value of X were displayed to both players. All subjects were informed prior to play that the value of X would be revealed to both players upon completion of the task, irrespective of the decisions made in the task. Revealing the value of X to all players ensures that Player As were not motivated to choose *In* (or *Conveying* X) just so Player A (Player B) could learn the value of X. This design feature eliminates any curiosity biases that may arise from the uncertainty regarding the value of X, and the consequent payoffs to other player.

After the subjects finished the game, they were asked to fill out a short questionnaire. In both treatments, the questionnaire contained 8 general demographic questions. In the PPT treatment, two additional questions were asked to both Player A and Player B that related to guilt feelings in the game. The specific questions that were asked to each Player, the corresponding responses, and a discussion of the possible insights that can be gleaned are presented in the results section.

4 Results

4.1 Aggregate Results and Hypothesis Testing

H1 stated that a larger proportion of Player As will *Convey* $X = 0$ compared to *Convey* $X = 6$ in Γ_{PPT}. To test this hypothesis, I consider Player As in the PPT Treatment who chose *In* and compare the conveyance decisions after $X = 0$ was revealed with the conveyance decision after $X = 6$ was revealed. Table 1 presents the aggregate Player A data from the PPT Treatment. From Table 1 we can see that of the 111 Player As, 58/111 (52%) chose *In.* Of those 58 Player As, 16 had $X = 0$ revealed and 26 had $X = 6$ revealed. In terms of the conveyance decision, 7/16 (44%) Player As chose to *Convey* $X = 0$ and 11/26 (42%) Player As chose to *Convey* $X = 6$. The proportion of Player As who *Convey* $X = 0$ is not significantly larger than the proportion who *Convey* $X = 6$ using a 1-sided Fisher's exact test ($p = 0.589$). The conveyance data suggests that Player As are not attempting to strategically induce guilt upon Player Bs, in the sense that there is no difference in

[16] I thank Urs Fischbacher (Fischbacher 2007) for providing this software.

the aggregate proportions of Player As who *Convey* $X = 0$ and *Convey* $X = 6$, which fails to support H1.

Table 1: Comparison of Conveyance Rates from PPT Treatment

Player A Data (PPT Treatment)			
In Rate	*Convey* $X = 0$	*Convey* $X = 6$	(p-value)[a]
58/111 (52%)	7/16 (44%)	11/26 (42%)	(0.589)

a: p-value is reported for a 1-sided Fisher's Exact test of conveyance rates

Although the aggregate data suggests that Player As are not attempting to induce guilt upon Players Bs, the data still allows me test H2a and H2b, albeit counterfactually. Despite the fact that there was no difference in the proportions of Player As who *Convey* $X = 0$ and *Convey* $X = 6$, we do observe Player As who *Convey* $X = 0$ and *Convey* $X = 6$. We can think of H2a and H2b as testing whether Player Bs are susceptible to strategic guilt induction, or alternatively, whether strategic guilt induction by Player A *would have been* effective at motivating Player B to choose *Right*. To test H2a and H2b, I compare the proportion of Player Bs choosing *Right* after $X = 0$ was conveyed (Right$|X = 0$), after the value of X was not conveyed (Right$|X = NC$), and after $X = 6$ was conveyed (Right$|X = 6$).

Table 2 presents the aggregate Player B data from the PPT Treatment. From Table 2, we can see that 3/7 (43%) Player Bs chose *Right*$|X = 0$, 5/40 (13%) chose *Right*$|X = NC$, and 0/11 (00%) choose *Right*$|X = 6$. A 1-sided Fisher's Exact test reveals a marginally significant difference between the proportion who chose *Right*$|X = 0$ and *Right*$|X = NC$ ($p = 0.084$). Although, the difference in the proportion who chose *Right*$|X = NC$ and *Right*$|X = 6$ is in the hypothesized direction, it is not significant ($p = 0.280$). However, using a Jonckheere-Terpstra test for ordered alternatives, we can reject the null that these three proportions are equal in favor of the ordered alternative that the proportion choosing *Right*$|X = 0 \geq$ *Right*$|X = NC$ $\geq$ *Right*$|X = 6$, with a least one strict inequality ($p = 0.008$). The data suggests that guilt induction by Player A, in the form of conveying $X = 0$, would have been effective at motivating Player B to choose *Right,* which supports H2a.

Table 2: Comparison of Right Rates from PPT Treatment

Player B Data (PPT Treatment)				
$Right\|X = 0$	$Right\|X = NC$	(p-value)[a]	$Right\|X = 6$	(p-value)[b]
3/7 (43%)	5/40 (13%)	(0.084)	0/11 (00%)	(0.280)

a: p-value is reported for a 1-sided Fisher's Exact test comparing Right$|X = 0$ and Right$|X = NC$

b: p-value is reported for a 1-sided Fisher's Exact test comparing Right$|X = NC$ and Right$|X = 6$

H3 stated that if having an opportunity to induce guilt fosters more trusting behavior, then more Player As would choose *In* when playing Γ_{PPT} compared to Γ_{UPT}. To test this hypothesis, I compare the aggregate Player A *In* rate in the PPT Treatment with the UPT Treatment. From Table 3, we can see that 58/111 (52%) of Player As chose *In* in the PPT Treatment and 17/45 (38%) of Player As chose *In* in the UPT treatment, which is marginally significant using a 1-sided Fisher's Exact test ($p = 0.071$). The data suggests that strategic settings that provide an opportunity for agents to induce guilt may foster more trusting behavior by those agents.

Table 3: Comparison of Player A In Rates

	UPT Treatment	PPT Treatment	(p-value)[a]
In Rate	17/45 (38%)	58/111 (52%)	(0.071)

a: p-value is reported from a 1-sided Fisher's Exact test of conveyance rates

4.2 Questionnaire Results

In this section I present results from two questions of a post decision questionnaire, related to perceptions of guilty feelings, that were asked to subjects in the PPT Treatment. For the analysis of the questionnaire data, I compare the responses of the two questions using matched samples based on the player role and the players' decisions. The questionnaire was not incentivized and did not impact monetary earnings. The motivation of the questionnaire was to gain insights regarding the players feelings, and perceptions of feelings, regarding guilt in Γ_{PPT}.

The two questions relating to perceptions of guilt were specific to the player role and the decisions made in the game. For Player As, the first question asked how much guilt they thought Player B felt (would have felt) from choosing *Left* if Player B knew (would have known) the true value of X. The second question asked Player As how much guilt they thought Player B felt (would have felt) from choosing *Left* if Player B did not know (would not have known) the true value of X. Responses were ranked on a 5-point scale with 5 being a *Very High* amount and 1 being a *Very Low* amount. Table 4 presents the aggregate response data from the two questions for Player As.[17]

Table 4: Player As Reported Player B Guilt Feeling

Panel A – Player As who chose Out (49 Subjects)

$X = 0$ known to Player B	$X = 0$ Unknown to Player B	$X = 6$ Known to Player B	$X = 6$ Unknown to Player B
2.90***	2.03	1.60	2.05*

Panel B – Player As who chose In (52 Subjects)

$X = 0$ Known to Player B	$X = 0$ Unknown to Player B	$X = 6$ Known to Player B	$X = 6$ Unknown to Player B
2.27**	1.64	1.50	2.50***

Notes: The reported amounts of guilt were compared using a 1-sided Wilcoxon Signed-Rank test for a matched sample.
* significance at 10% level ** significance at 5% level *** significance at 1% level

The Player A response data in Table 4 is divided into 2 panels corresponding to the *In/Out* decision made by Player A. Each of the panels is then further divided

[17]Because the questionnaire was voluntary, subjects were not required to submit an answer for each question. A total of 10 of 111 Player As did not answer at least one of the questions related to there belief of Player Bs guilt feelings. Therefore, the aggregate data in Table 3 reflects the responses of 101 Player As who did answer both questions.

based on the true value of X. From Panel A, we see that Player As who chose *Out* perceived that Player B would have felt more guilt from choosing *Left* if $X = 0$ was conveyed compared to if $X = 0$ was not conveyed (p = 0.001), and more guilt if $X = 6$ was not conveyed compared to if $X = 6$ was conveyed (p = 0.082). Similarly, From Panel B, we see that Player As who chose *In* perceived that Player B would have felt more guilt from choosing *Left* if $X = 0$ was conveyed compared to if $X = 0$ was not conveyed (p = 0.017), and more guilt if $X = 6$ was not conveyed compared to if $X = 6$ was conveyed (p = 0.0002). The questionnaire response data provides evidence that Player As recognized, at least ex-post, that guilt *could have been* induced in Player B by conveying $X = 0$ and not conveying $X = 6$, despite behavior that is inconsistent with Player As attempting to do so.

For Player B, the first question asked to rate the amount of guilt he/she felt (would have felt) from choosing *Left* (if he had chosen *Left*) if he/she did not know the value of X. The second question asked Player B to rate the amount of guilt he/she felt (would have felt) from choosing *Left* (if he had chosen *Left*) if he/she knew the value of X. Again, responses were ranked on a 5-point scale with 5 being a *Very High* amount and 1 being a *Very Low* amount. Table 5 presents the aggregate response data from the two questions for Player Bs.[18]

Table 5: Player Bs Reported Guilt Feelings

Panel A – Player Bs who did not make a decision (52 Subjects)			
$X = 0$ Known to Player B	$X = 0$ Unknown to Player B	$X = 6$ Known to Player B	$X = 6$ Unknown to Player B
2.32***	1.55	1.57	2.29***
Panel B – Player Bs who made a decision (53 Subjects)			
$X = 0$ Known to Player B	$X = 0$ Unknown to Player B	$X = 6$ Known to Player B	$X = 6$ Unknown to Player B
2.22***	1.61	1.60	2.20***

Notes: The reported amounts of guilt were tested using a 1-sided Wilcoxon Signed-Rank test for a matched sample.
* significance at 10% level ** significance at 5% level *** significance at 1% level

[18] A total of 6 of 111 Player Bs did not answer at least one of the questions related to their guilt feelings. Therefore, the aggregate data in Table 4 reflects the responses of 105 Player Bs who did answer both questions.

The Player B response data in Table 5 is divided into 2 panels that correspond to whether Player B actually made a decision, i.e., whether Player A chose *In.* From Panel A we see that Player Bs who did not make a decision responded that they would have felt more guilt from choosing *Left* if $X = 0$ was conveyed compared to if X was not conveyed (p = 0.0001), and more guilt from choosing *Left* if X was not conveyed compared to if $X = 6$ was conveyed (p = 0.0001). Similarly, the Player Bs who made a decision responded that they did (would have) felt more guilt from choosing *Left* if $X = 0$ was conveyed compared to if X was not conveyed (p = 0.003), and more guilt from choosing left if X was not conveyed compared to if $X = 6$ was conveyed (p = 0.004). The questionnaire response data from Player Bs reveals that Player Bs, irrespective of whether or not they made a decision, were susceptible to guilt induction by Player A, which is consistent with the observed Player B behavior.

4.3 Additional Results

In this section I present the results from an auxiliary treatment that was run to complement the results from the PPT Treatment. The third treatment, which I call the PPD Treatment, involved subjects playing a "dictator" version of Γ_{PPT}. In this modified dictator version of Γ_{PPT}, the initial *In/Out* decision of Player As was eliminated. As a result, Player As only face the decision of whether they want to convey the value of X to Player B, condition on the value of X being revealed. Then, conditional on the conveyance state, Player B decides whether to choose *Right* or *Left.* I refer to this dictator version of Γ_{PPT} as Γ_{PPD}. The extensive form of Γ_{PPD} is simply the subgame of Γ_{PPT} that begins with Nature's move.

The motivation for considering this additional PPD Treatment is twofold. First, Γ_{PPD} enables Player As to attempt to induce guilt in the same manner as Γ_{PPT} by choosing to *Convey* $X = 0$ and *Not Convey* $X = 6$. Thus, the PPD Treatment allows me to test H1 in a more simplified strategic setting, which serves as a robustness test of Player As failure to attempt to induce guilt in the PPT Treatment. Second, Γ_{PPD} provides an opportunity to test for the effectiveness of guilt induction minus concerns for reciprocity by Player B that might be present in Γ_{PPT} (see Cox 2004 for a discussion of reciprocity motivations in trust games).[19] Thus, the PPD Treatment allows me to test H2a and H2b while eliminating any possible reciprocity confounds.

Table 6 presents the Player A conveyance data from the PPD Treatment. From Table 6, we can see that 23/27 (85%) Player As conveyed $X = 0$ and 22/29 (75%) Player As conveyed $X = 6$, which is not significant using a 1-sided Fisher's Exact

[19] The general idea behind preferences for reciprocity is that agents are motivated to respond kindly to agent who are kind to them, and respond unkindly to agents who are unkind to them (see Dufwenberg and Kirchsteiger 2004 for a formal model of reciprocity). In the context of Γ_{PPT}, Player B may be motivated to reciprocate the kind action of Player A choosing *In* by choosing *Right.* This motivation may be more prevelant if Player B knows that $X = 0$. However, Γ_{PPD} avoids such a possibility.

test ($p = 0.296$). Although a larger proportion of Player As convey $X = 0$ in the PPD Treatment, relative to the PPT Treatment, I still do not find evidence that supports H1 due to the larger proportion of Player As who convey $X = 6$ in the PPD Treatment.

Table 6: Comparison of Conveyance Rate From PPD Treatment

Player A Data (PPD Treatment)		
Convey $X = 0$	*Convey* $X = 6$	(p-value)[a]
23/27 (85%)	22/29 (75%)	(0.296)

a: p-value is reported from a 1-sided Fisher's Exact test of conveyance rates

Table 7: Comparison of Right Rates From PPD Treatment

Player B Data (PPD Treatment)				
$Right \mid X = 0$	$Right \mid X = NC$	(p-value)[a]	$Right \mid X = 6$	(p-value)[b]
9/23 (39%)	2/21 (10%)	(0.026)	2/22 (9%)	(0.679)

a: p-value is reported for a 1-sided Fisher's Exact test comparing Right$|X = 0$ and Right$|X = NC$

b: p-value is reported for a 1-sided Fisher's Exact test comparing Right$|X = NC$ and Right$|X = 6$

Table 7 presents the aggregate Player B data from the PPD Treatment. From Table 2, we can see that 9/23 (39%) of Player Bs chose $Right|X = 0$, 2/21 (10%) chose $Right|X = NC$, and 2/22 (9%) choose $Right|X = 6$. These proportions are very similar to those observed in the PPT Treatment. Again, a 1-sided Fisher's Exact test reveals a significant difference between the proportion who chose $Right|X = 0$ and

Right$|X = NC$ ($p = 0.026$). Although, again, a 1-sided Fisher's Exact test reveals no significant difference in the proportion who chose *Right*$|X = NC$ and *Right*$|X = 6$ ($p = 0.679$). A Jonckheere-Terpstra test for ordered alternatives does reject the null that these three proportions are equal in favor of the ordered alternative that the proportion choosing *Right*$|X = 0 \geq$ *Right*$|X = NC \geq$ *Right*$|X = 6$, with a least one strict inequality ($p = 0.006$). This data confirms that results from the PPT Treatment that guilt induction by Player A, in the form of conveying $X = 0$, would have been effective at motivating Player B to choose *Right,* which provides further support H2a.

4.4 Discussion

Although the motivation of this study was to test H1-H3, I take this time to make a few speculative remarks regarding the observed data that is inconsistent with H1 and H2b. I begin by considering the inconsistency with H1. Recall that attempted strategic guilt induction by Player A in Γ_{PPT} (and Γ_{PPD}) requires Player A convey $X = 0$ and not convey $X = 6$. Thus, failure to attempt to induce guilt can result from Player As either (i) not conveying $X = 0$, (ii) conveying $X = 6$, or (iii) both.

For the most part, a large percentage of Player As do convey $X = 0$ (70% when aggregated over the PPT and PPD Treatments). However, a large percentage of Player As also convey $X = 6$ (60% when aggregated over the PPT and PPD Treatments), which suggests that failure to support H1 is primarily a result of (ii). There are several possible explanations why Player A might be motivated to convey $X = 6$ in Γ_{PPT} and Γ_{PPD}. One possibility is that Player As may have some aversion to being deceptive. Player As may perceive failing to convey information as being deceptive and, as a result, are motivated to convey X, regardless of the value of X. A second possibility is that Player As are inequality averse and prefer the (6,6) outcome to the (10,4) outcome.[20] If Player A prefers (6,6) to (10,4), then Player A would be motivated to *Convey* $X = 6$ in an attempt to persuade Player B to choose *Left,* i.e., dissuade Player B from choosing *Right.*

BSH (1994) posit that guilt is induced by conveying how one suffers if another fails to act in a desired fashion. You can think of Player A choosing to convey $X = 0$ as an explicit attempt to induce guilt because Player A is actually conveying how he suffers, as posited by BSH (1994). However, you can think of Player A choosing to not convey $X = 6$ as an implicit attempt to induce guilt because Player A is convey how he suffers by actually not conveying how he doesn't suffer. It is possible that Player As simply don't conceptualize this implicit method for inducing guilt by not conveying how you don't suffer.

[20] According to the Fehr and Schmidt (1999) model, Player A would prefer the payoff vector of (6,6) over (10,4) if $6 \geq 10 - \beta(10 - 4)$. Therefore, a Player A who is averse enough to inequality, ($\beta \geq 2/3$), would prefer the (6,6) payoff vector to the (10,4) payoff vector.

In terms of this distinction between explicit and implicit guilt induction, we can think of H2a as testing whether Player Bs are susceptible to explicit guilt induction and H2b as testing whether Player Bs are susceptible to implicit guilt induction. Given that the data supports H2a and fails to support H2b, it may be the case that Player Bs do not conceptualize that Player As suffer more from their choice of *Left* when X is not conveyed, compared to when $X = 6$ is conveyed. This suggests that guilt induction can be effective, but perhaps only when agents explicitly convey how they suffer. That is, if an agent knows he will suffer over another's failure to act in a desired fashion, then that agent may be able to induce guilt by explicitly conveying that information.

5 B&D Model of Guilt Applied to Γ_{PPT}

In this section, I explore the connection between the psychological insights of BSH (1994) regarding guilt induction and the B&D model of guilt. In doing so, I first provide a brief outline of the general B&D model of guilt, followed by its application to Γ_{PPT}. I then derive conditions under which the BSH (1994) method of inducing guilt is consistent with predictions of the B&D model of guilt. I conclude the section by showing that guilt induction, as prescribed by BSH (1994), can be supported as an equilibrium in Γ_{PPT} under the framework of B&D.

5.1 Outline of B&D Model of Guilt

B&D formally model two types of guilt for a general class of extensive form games, simple guilt and guilt from blame. I only consider the former applied to Γ_{PPT}, thus, I restrict my analysis to simple guilt.[21] What follows is only a simplified outline of the model; Interested readers should refer to B&D for the full, technical presentation of both models including illustrative examples. Informally, the model posits that agents suffer disutility, in the form of guilt, from failing to live up to others' expectations. This is captured by modeling an agents utility as a function of his own material payoffs, and the extent to which he let other agents down.

Formally, simple guilt is modeled by specifying a utility function for player i given by:

$$u_i^{SG} = m_i - \sum_{j \neq i} \theta_{ij} \cdot D_j \qquad \text{(Simple Guilt Utility)}$$

[21] With simple guilt, an agent suffers disutility proportional to how much he/she lets down another agent. Whereas with guilt from blame, an agent suffer disutility proportional to how much the other agent blames him for being let down. Thus, the main difference between the two models is the extent to which an agent can be blamed for letting down another agent. With respect to Γ_{PP}, the two models are equivalent. Player A can unambiguously identify the action of Player B which implies that Player B will receive all the blame for letting Player A down. Although either model of guilt could be used for this analysis, I opt to apply the less complex model of simple guilt for clarity.

In this expression, m_i represents player $i\prime s$ material payoff, and $\sum_{j\neq i}\theta_{ij}\cdot D_j$ represents player $i\prime s$ disutility from simple guilt. The latter component is composed of two pieces. The first, θ_{ij}, is an exogenously given constant that measures player $i\prime s$ sensitivity of feeling guilty toward player j. The second, D_j, represents the amount by which player i lets player j down, as a result of player $i\prime s$ strategy. $D_j = E_j - m_j$ is expressed as the difference between the material payoff that player j was expecting, E_j, and the material payoff that player j actually receives, m_j.[22] E_j itself is a function of player $j\prime s$ strategy, and player $j\prime s$ vector of "first-order" beliefs regarding the strategies of the other players. Note, player i does not actually observe D_j, as it is a function of player $j\prime s$ first-order beliefs. Therefore, it is assumed that player i maximizes the expected value of u_i^{SG}, given player $i\prime s$ first-order beliefs regarding player $j\prime s$ strategy, and player $i\prime s$ "second-order" belief regarding player $i\prime s$ first-order beliefs.

5.2 Guilt Applied to Γ_{PPT}

I proceed with the application of the B&D model of simple guilt to Γ_{PPT}.[23] In particular, I derive the guilt that Player B would experience from choosing *Left* in Γ_{PPT}. When Player B is called upon to move, there are three possible histories that Player B could observe. The first is that $X = 0$ was conveyed, which I denote C^0. The second is that $X = 6$ was conveyed, which I denote C^6. The third is that the value of X was not conveyed, which I denote C^N. Henceforth, I refer to these three possible histories as the three possible conveyance states. A strategy for Player B is a probability distribution over Player B's possible actions, {*Left, Right*}, at each of the three possible conveyance states.

In deriving the guilt that Player B would suffer from choosing *Left*, it is necessary to first derive Player A's material expectation. Before Player A makes her initial *In* or *Out* decision, Player A forms an initial first-order belief regarding Player B's strategy. Player A's initial first-order beliefs can be represented as the probabilities that Player B would choose *Right* at each of the three conveyance states, which I denote $\boldsymbol{\alpha}_A = (\Pr(Right|C^0), \Pr(Right|C^N), \Pr(Right|C^6))$. For a given strategy, Player A forms an expectation,weighted over $\boldsymbol{\alpha}_A$ and moves by Nature, of his material payoff. I denote Player A's initial material expectation E_A.

Player B experiences disutility from simple guilt when he fails to live up to Player

[22] More precisely, B&D model the utility from simple guilt as $u_i^{SG} = m_i - \sum_{j\neq i}\theta_{ij}\cdot G_{ij}$. In this expression, $G_{ij} = D_j - \min_{s_i} D_j$ measures the unavoidable amount by which player i lets down player j. However, in relation to Γ_{PP}, G_{ij} is equivalent to D_j so I opt to use the more simple notation.

[23] The guilt aversion model could be applied to Γ_{UP} as well. However, the hypotheses that I develop regarding attempted guilt induction and the effectiveness of guilt induction apply only to Γ_{PP}. Therefore, I only consider the guilt aversion model applied to Γ_{PP}. Γ_{UP} is included in the experimental design solely to investigate whether having an opportunity to induce guilt affects trusting behavior as described in the previous Section.

A's expectations i.e. chooses a strategy that yields a payoff to Player A that is lower than E_A. However, when Player B is called upon to move, he does not observe E_A. Player B must then form an expectation regarding E_A, conditional on the conveyance state. This conditional expectation is a function of Player B's conditional second-order beliefs regarding $\boldsymbol{\alpha}_A$, which I denote $\boldsymbol{\beta}_B$. I define $E_B|h \; where \; h \in \{C^0, C^N, C^6\}$ as Player B's expectation of E_A, conditional on the conveyance state. The guilt that Player B would suffer from choosing *Left* is proportional to the difference between $E_B|h$ and m_A, where m_A the material payoff that Player A actually receives as a result of Player B's *Left* decision. Let $\theta_B \geq 0$ denote Player B's sensitivity to feeling guilty. Below, I derive Player B's utility, including simple guilt, from choosing *Left* at each of the three possible conveyance states.

X = 0 was conveyed: Because the conveyance is credible, Player B knows that if he chooses *Left,* Player A will receive a payoff of $m_A = 0$. Player B's expectation of Player A's expectation is $E_B|C^0$. By choosing *Left*, Player B will suffer disutility from guilt equal to $\theta_B \cdot (E_B|C^0 - 0)$. Thus, Player B's utility from choosing *Left*, after $X = 0$ was conveyed, is equal to:

$$6 - \theta_B \cdot (E_B|C^0 - 0)$$

X = 6 was conveyed: Because the conveyance is credible, Player B knows that if he chooses *Left,* Player A will receive a payoff of $m_A = 6$. Player B's expectation of Player A's expectation is $E_B|C^6$. By choosing *Left,* Player B will suffer disutility from guilt equal to $\theta_B \cdot (E_B|C^6 - 6)$. Thus, Player B's utility from choosing *Left*, after $X = 0$ was conveyed, is equal to:

$$6 - \theta_B \cdot (E_B|C^6 - 6)$$

Value of X not conveyed If the value of X is not conveyed, then Player B must think about the expected material payoff that Player A would receive if he chooses *Left.* Let $\widehat{m}_A = E_B[m_A|C^N]$ denote this expectation. In calculating $\widehat{m}_A$, Player B must think about the relative probability that (1) Player A learned $X = 0$ and chose *Not Convey,* (2) Player A learned $X = 6$ and chose *Not Convey*, and (3) Player A did not learn the value of X. Although these probabilities are unobservable to the researcher and, thus, the exact value of $\widehat{m}_A$ is unobservable, it is possible to identify bounds on $\widehat{m}_A$. Specifically, $\widehat{m}_A \in [1, 5]$ as I previously shown. Player B's expectation of Player A's expectation is $E_B|C^N$. By choosing *Left*, Player B will suffer disutility from guilt equal to: $\theta \cdot (E_B|C^N - \widehat{m}_A)$ where $\widehat{m}_A \in [1, 5]$. Thus, Player B's utility from choosing *Left*, after the value of X was not conveyed, is equal to:

$$6 - \theta_B \cdot (E_B|C^N - \widehat{m}_A) \; where \; \widehat{m}_A \in [1, 5]$$

5.3 Theoretical Consistency with BSH (1994)

BSH (1994) argue that guilt is induced by conveying to another how one suffers when the other fails to act in a desired fashion. In relation to Γ_{PPT}, the BSH (1994) method of inducing guilt would correspond to Player A choosing to *Convey* $X = 0$, and choosing to *Not Convey* $X = 6$. Is the BSH (1994) method of guilt induction consistent with the B&D model of simple guilt? That is, does the B&D model predict that Player B would suffer more guilt from choosing *Left* when $X = 0$ was conveyed compared to when the value of X was not conveyed? Similarly, does the model predict that Player B would suffer more guilt from choosing *Left* when the value of X was not conveyed compared to when $X = 6$ was conveyed? These questions can both be conditionally answered in the affirmative. I proceed by deriving conditions under which the BSH (1994) method for inducing guilt is consistent with the prediction of the B&D model of simple guilt.

First recall that the disutility from guilt that Player B would suffer from choosing *Left* at each of the three possible conveyance states, as predicted by the B&D model, can be summarized as follows:

- $X = 0$ was conveyed: $\theta_B \cdot (E_B|C^0 - 0)$
- $X = 6$ was conveyed: $\theta_B \cdot (E_B|C^6 - 6)$
- Value of X not conveyed: $\theta_B \cdot (E_B|C^N - \widehat{m}_A)$ where $\widehat{m}_A \in [1, 5]$

According to B&D, Player B would suffer more disutility from choosing *Left* when $X = 0$ was conveyed, compared to when the value of X was not conveyed, when:

$$\theta_B \cdot (E_B|C^0 - 0) \geq \theta_B \cdot (E_B|C^N - \widehat{m}_A) \Longrightarrow E_B|C^0 \geq E_B|C^N - \widehat{m}_A \quad \text{(Condition \#1)}$$

Similarly, Player B would suffer more disutility from choosing *Left* when the value of X was not conveyed, compared to when $X = 6$ was conveyed, when:

$$\theta_B \cdot (E_B|C^N - \widehat{m}_A) \geq \theta_B \cdot (E_B|C^6 - 6) \Longrightarrow E_B|C^N - \widehat{m}_A \geq E_B|C^6 - 6 \quad \text{(Condition \#2)}$$

Because $\widehat{m}_A \in [1, 5]$, we can derive a set of necessary and sufficient conditions on $E_B|h$ under which guilt induction, as prescribed by BSH (1995), is consistent with the prediction of the B&D model of simple guilt. The two conditions that are necessary for guilt induction by Player A upon Player B in Γ_{PPT} are:

$$E_B|C^0 \geq E_B|C^N - 5 \quad \text{(Necessary Condition \#1)}$$

$$E_B|C^N \geq E_B|C^6 - 5 \quad \text{(Necessary Condition \#2)}$$

The two conditions that are sufficient for guilt induction by Player A upon Player B in Γ_{PPT} are:

$$E_B|C^0 \geq E_B|C^N - 1 \qquad \text{(Sufficent Condition \#1)}$$

$$E_B|C^N \geq E_B|C^6 - 1 \qquad \text{(Sufficent Condition \#2)}$$

Essentially, Condition 1 states that Player B's second-order belief of Player A's expectation after $X = 0$ is conveyed, $E_B|C^0$, is not *too* much lower than Player B's second-order belief of Player A's expectation after the value of X is not conveyed, $E_B|C^N$. Similarly, Condition 2 states that Player B's second-order belief of Player A's expectation after the value of X is not conveyed, $E_B|C^N$, is not *too* much lower than Player B's second-order belief of Player A's expectation after $X = 6$ is conveyed, $E_B|C^6$. Conditions 1 and 2 would certainly be satisfied if we assumed that Player B does not update his belief of Player A's expectation, i.e. $E_B|C^0 = E_B|C^N = E_B|C^6$. Such an assumption would be satisfied in an equilibrium as I will discuss in the subsequent section. However assuming that $E_B|C^0 = E_B|C^N = E_B|C^6$ is clearly stronger than is needed for Condition 1 and 2 to be satisfied.[24]

If Conditions 1 and 2 are satisfied, then the B&D model predicts that Player B would feel more guilt from choosing left after $X = 0$ was conveyed and the value of X was not conveyed, compared to when the value of X was not conveyed and $X = 6$ was conveyed, respectively. Therefore, if Condition 1 and 2 hold, then according to the B&D model, Player A can induce guilt upon Player B by choosing to *Convey* $X = 0$ and *Not Convey* $X = 6$, which is consistent with the BSH (1994) method for inducing guilt.

5.4 Guilt Induction as an Equilibrium of Γ_{PPT}

The primary motivation of the study is to experimentally investigate whether agents strategically induce guilt upon others, and whether agents respond in kind to strategic guilt induction. These are questions related to behavioral motivations in games that are independent of any equilibrium supposition. Nevertheless, it is an important question whether such behavior can be supported as an equilibrium of Γ_{PPT} under the formal guilt framework of B&D. It is the case that guilt induction by Player A, and a kind response to guilt induction by Player B, can be supported as sequential equilibrium (SE) of Γ_{PPT} (Battigalli and Dufwenberg 2009).[25] The intuition behind

[24] Conditions 1 and 2 are a function of both Player B's updated second-order belief of Player A expectation, and Player B's belief about the value of X when the value of X is not conveyed. Although these conditions can't be empirically verified for each individual Player B, it is possible to verify these conditions on the aggregate level.

[25] I refer the reader to the original paper for a formal equilibrium analysis of *dynamic psychological games.* The authors extend the concept of sequential equilibrium by incorporating hierarchies of conditional beliefs.

this rests in the fact that in a SE, an assessment (profile of behavioral strategies and conditional hierarchical beliefs) will be consistent. Battigalli and Dufwenberg show that in equilibrium, "players never change their beliefs about the conditional beliefs that the opponents would hold at each h (history)" (pp. 16). Thus, in equilibrium, there is no belief updating. It follows that $E_B|C^0 = E_B|C^N = E_B|C^6$, which implies that Conditions 1 and 2 from above would be satisfies in an equilibrium.

As I have shown, guilt induction by Player A is characterized by the choice to *Convey* $X = 0$ and *Not Covey* $X = 6$. Therefore, the strategy (*In, Convey* $X = 0$, *Not Convey* $X = 6$) for Player A is consistent with attempted guilt induction. Effective guilt induction is characterized by Player B choosing *Right* as a response to the guilt induction by Player A. The following two strategies for Player B are consistent with responding in kind to Player A's guilt induction: (*Right*$|C^0$, *Right*$|C^N$, *Left*$|C^6$), and (*Right*$|C^0$, *Left*$|C^N$, *Left*$|C^6$). Thus, the strategy profiles ((*In, Convey* $X = 0$, *Not Convey* $X = 6$), (*Right*$|C^0$, *Right*$|C^N$, *Left*$|C^6$)), and ((*In, Convey* $X = 0$, *Not Convey* $X = 6$), (*Right*$|C^0$, *Left*$|C^N$, *Left*$|C^6$)) are the candidate equilibrium profiles for effective guilt induction in Γ_{PPT}. I proceed by showing that each can be supported as a SE of Γ_{PPT}.

Claim 1 The strategy profile ((*In, Convey* $X = 0$, *Not Convey* $X = 6$), (*Right*$|C^0$, *Right*$|C^N$, *Left*$|C^6$)) can be supported as a SE of Γ_{PPT} for $\theta_B \in [\frac{2}{5}, \frac{1}{2}]$

To verify that this strategy profile is an equilibrium, we need to check that neither player has a profitable deviation. For Player A, this is rather trivial. By following the equilibrium strategy, Player A earns a payoff of 10, which is the highest payoff of the game. Therefore, Player A has no profitable deviation. For Player B, we need to consider deviations at each of the possible conveyance states. Given consistent beliefs in equilibrium, we have that $\boldsymbol{\alpha}_A = \boldsymbol{\beta}_B = (1, 1, 0)$, $E_A = E_B|h = 10 \; \forall \; h \in \{C^0, C^N, C^6\}$, and $\widehat{m}_A = 5$. Player B will not deviate to *Right*$|C^6$ so long as: $6 - \theta_B \cdot [10 - 6] \geq 4 \Leftrightarrow \theta_B \leq \frac{1}{2}$. Player B will not deviate to *Left*$|C^N$ so long as: $4 \geq 6 - \theta_B \cdot [10 - 5] \Leftrightarrow \theta_B \geq \frac{2}{5}$. Similarly, Player B will not deviate to *Left*$|C^0$ so long as: $4 \geq 6 - \theta_B \cdot [10 - 0] \Leftrightarrow \theta_B \geq \frac{1}{5}$ which is satisfied if $\theta_B \geq \frac{2}{5}$.

Claim 2 The strategy profile ((*In, Convey* $X = 0$, *Not Convey* $X = 6$), (*Right*$|C^0$, *Left*$|C^N$, *Left*$|C^6$)) can be supported as a SE of Γ_{PPT} for $\theta_B \in [\frac{2}{7}, 1]$

Again, to verify that this strategy profile is an equilibrium, we need to check that neither player has a profitable deviation. For Player A, playing the equilibrium strategy yields an expected payoff of 6.75, therefore Player A cannot profitable deviate to *Out.* Player A would not deviate and *Not Convey* $X = 0$ which would result in a payoff of 0 compared to a payoff of 10 from following the equilibrium strategy to *Convey* $X = 0$. Player A is indifferent between *Convey* $X = 6$ and *Not Convey* $X = 6$. Therefore, Player A has no profitable deviation from the prescribed equilibrium strategy. For Player B, we need to consider deviations at each of the possible conveyance states. Given consistent beliefs in equilibrium, we have that $\boldsymbol{\alpha}_A = \boldsymbol{\beta}_B = (1,0,0)$, $E_A = E_B|h = 7 \ \forall \ h \in \{C^0, C^N, C^6\}$, and $\widehat{m}_A = 5$. Player B will not deviate to *Right*$|C^6$ so long as: $6 - \theta_B \cdot [7-6] \geq 4 \Leftrightarrow \theta_B \leq 2$. Player B will not deviate to *Right*$|C^N$ so long as: $6 - \theta_B \cdot [7-5] \geq 4 \Leftrightarrow \theta_B \leq 1$. Similarly, Player B will not deviate to *Left*$|C^0$ so long as: $4 \geq 6 - \theta_B \cdot [7-0] \Leftrightarrow \theta_B \geq \frac{2}{7}$.

Claim 1 and 2 show that strategic guilt induction can be supported as an equilibrium of Γ_{PPT} under the formal guilt framework of B&D. I acknowledge that an equilibrium supposition, especially when a game features multiple equilibrium, is a rather strong notion. However, an equilibrium supposition is sufficient, and not necessary, for the research hypotheses of this study to be consistent with predictions of the B&D model of guilt.

6 Conclusion

The motivation of this study is to investigate some of the interpersonal strategic implications from guilt aversion. In doing so, I consider an experimental design centered around a private information trust game – Γ_{PPT}, which features a strategic structure rich enough to allow agents an opportunity to induce guilt in a manner consistent with psychological insights of BSH (1994). I then test whether agents strategically induce guilt upon others, whether strategic guilt induction is effective, and whether having the opportunity to induce guilt fosters trust.

In general, the results fail to support the hypothesis (H1) that agents attempt to strategically induce guilt upon others. Although there is some evidence that Player As (first movers) explicitly convey to Player B how much they would suffer, which is consistent with BSH (1994), Player As also convey how much they do not suffer. That is, Player As tend to pool on conveying payoff relevant information to Player B. However, the data does support the hypothesis (H2a) that Player Bs are more likely to be kind to Player A after Player A has explicitly conveyed how much he would suffer if Player B acts unkindly. That is, Player Bs appear to be susceptible to strategic guilt induction. Additionally, the data marginally supports the hypothesis (H3) that Player As exhibit more trusting behavior when the strategic setting features and opportunity to induce guilt upon Player B.

The observed data that Player Bs are susceptible to guilt induction provides additional experimental evidence consistent with the hypothesis that agents are motivated by guilt aversion. Thus, Γ_{PPT} and Γ_{PPD} allow for identification of behavior that is consistent with guilt aversion without having to elicit beliefs. The experimental design provided an alternative approach for investigating guilt aversion from those previously implemented that does not require belief elicitation or conveying elicited beliefs, both of which present previously established limitations (c.f. Charness and Dufwenberg

2011).[26] The ability to test for guilt aversion without eliciting beliefs is particularly relevant in light of the recent studies by Reuben, Sapienza, and Zingales (2009) and Ellingsen, Johannesson, Tjotta, and Torsvik (2010), which both test for the presence of guilt aversion using similar experimental designs that feature belief elicitation, yet reach opposing conclusions. This paper joins Charness and Dufwenberg (2011), in its ability to test models of belief dependent utility without having to elicit beliefs.

BSH(1994) note that "guilt [induction] does not depend on formal power or influence and may even work best in the absence of such power, because one induces guilt by depicting oneself as the helpless victim of another's actions" (p. 247). This suggests that guilt induction could be particularly effective in economies with less developed legal systems. In such economies guilt induction could serve as an informal mechanism for enforcing contracts and mitigating corrupt behavior, that might otherwise transpire in the absence of formal prohibitive legislation (Lee 2010). Guilt induction could also prove to be effective at influencing behavior and impacting outcomes in credence goods markets (see Dulleck and Kerschbamer 2006; Dulleck, Kerschbamer, and Sutter 2009; Beck, Kerschbamer, Qiu, and Sutter 2010). In these credence goods markets, e.g. doctors, mechanics, or other expert services, the consumer is often the "helpless victim" of the experts actions. Guilt induction by the consumer could be implemented to thwart opportunistic behavior by the expert, especially in developing economies where the incentives for opportunistic behavior are likely to be much stronger.

Partnerships, principle-agent contracting, and employee-employer relationships, represent some of the many economic settings where trust is pivotal for successful and efficient relations. There is a growing body of literature that investigates the importance of trust in social and economic settings, and how trust can be fostered (see Fehr 2009; Charness, Du, Yang 2011 for reviews). Much of this literature focuses on the effectiveness of reputation building in fostering trust.[27] While there is often

[26] Specifically, Dufwenberg and Gneezy (2000), Charness and Dufwenberg (2006), Bacharach, Guerra, and Zizzo (2007), and Dufwenberg, Gächter, and Hennig-Schmidt (2011) elicited second order beliefs and test for a positive correlation between elicited second order expectations and actions. However, because these studies provide only a correlation between elicited second order beliefs and actions, anchoring and false consensus effects cannot be ruled out as possible explanations. Alternatively, Reuben, Sapienza, and Zingales (2009) and Ellingsen, Johannesson, Tjotta, and Torsvik (2010) elicit first-order expectations of subjects, convey those expectations to the subject's partner, and test for correlations between expectations and actions. However, the possibility of untruthful reporting of beliefs and skepticism of conveyed beliefs arise with this approach, as noted by Reuben, Sapienza, and Zingales.

[27] That is, building a trustworthy reputation through prior trustworthy actions, that are observable to other agents, induces agents to trust you in the future. Many experimental studies have found evidence consistent with this "indirect reciprocity" including Bohnet and Huck (2004), Bolton, Katok, and Ockenfels (2005), Greiner and Levati (2005), Seinen and Schram (2006), Duffy, Lee, and Xie (2008), and Engelmann and Fischbacher (2009). Charness, Du, and Yang (2010) provide a thorough review of these papers as well as provide experimental evidence that a reputation of trusting behavior can foster trust.

an incentive to trust in economic settings, this incentive is often offset by exposure to the risk of opportunistic behavior by the trusted agent. However, guilt induction by the trusting agent can serve as a mechanism for thwarting such opportunistic behavior, thus mitigating the risk associated with trusting actions. Therefore, having an opportunity to induce guilt would then lead to more trusting behavior, which is what is observed in this paper. This might help explain why trust is so prevalent in many economic interactions in our society today, where the strategic environments are often rich enough to allow the possibility to induce guilt.

Before I conclude, I speculate about two possible areas of future research that might prove valuable for better understanding the interpersonal implications of guilt aversion in strategic settings. The first involves investigating guilt induction in the absence of credible payoff conveyance between agents. The experimental game in this study only permits credible payoff conveyance between agents. However, in many economic setting with private payoff information, credible payoff conveyance may be costly or impossible, e.g., firms involved in relationship specific investment or partnerships. It remains to be investigated whether agents attempt to induce guilt, and whether induced guilt remains an effective behavior mechanism when credible payoff revelation is not possible.[28] The second involves investigating guilt induction and its effectiveness in repeated games. Specifically, ex post guilt induction upon an agent who chose an unkind action in a previous period could thwart unkind actions by that agent in future periods.[29] Many economic settings involve repeated interaction between agents and in these settings, guilt induction could be used to foster long lasting, trusting relationships.

I conclude by noting that the effectiveness of guilt induction as an influence mechanism in strategic settings may have limitations. In particular, repeated applications of guilt induction may become less effective since the target of the guilt induction will likely become resentful or angered over its repeated application. This could ultimately lead to less kind actions in response to guilt induction, which is counter to its intended purpose. BSH (1995) recognize this and argue that "although guilt may often be an effective way of getting one's way, it appears to be costly and to carry some stigma. This suggests that inducing guilt may be a technique that has to be used with caution and restraint" (p. 183). Perhaps guilt induction in strategic economic settings should be a mechanism that is reserved for instances when the payoff and/or risk associated with a trusting action are the largest.

[28] The credible conveyance aspect of the design played a critical role in the application of the B&D model of guilt aversion. Applying their model, deriving Player B's guilt, and showing that guilt induction can be supported as an equilibrium in a trust game with a "cheap talk" conveyance stage may prove to be non-trivial.

[29] In this type of setting, an agent would actually be inducing guilt, and not counterfactual guilt like I have considered in this study.

7 Appendix

Sample Instructions – PPT Treatment

Welcome and thank you for participating. Your participation is VOLUNTARY, and you may leave at any time. Feel free to raise your hand and ask questions at any time, and you may refer back to these instructions at any time during the session. Please remain seated and quiet for the remainder of the session. All decisions are to be completed individually and interaction with other participants is strictly PROHIBITED. Thank you for your cooperation.

Each person will receive a $5 show-up payment for participating. In addition, you can receive additional compensation based on the decision(s) that are made in the decision task described below. After the task is complete, you will be privately paid the amount of money you have earned. Upon completions of the decision task, please remain quietly seated in your carrel until you have been paid.

The Decision Task:

You will be participating in a 2-person decision task. Each person will be randomly and anonymously paired with another person in the lab. In each of the 2-person decision-making pairs, one person will be randomly assigned the role of PLAYER A and the other person will be randomly assigned the role of PLAYER B. You will remain in your assigned role for the entire session. The earnings of each Player will depend on the decision(s) he/she makes, and/or the decision(s) of the Player with whom they are paired. A brief outline of steps of the decision task will first be provided, followed by a detailed description of each step and the corresponding earnings for each Player.

- Step 1: PLAYER A begins by first choosing IN or OUT.
  - If PLAYER A chooses OUT, then the task ends.
  - If PLAYER A chooses IN, then the task proceeds to Step 2.
- Step 2: PLAYER A might privately learn some payoff information that was initially unknown to both players. If PLAYER A does learn the information, the PLAYER A will then have an opportunity to convey the information to PLAYER B. The task will then proceed to Step 3.
- Step 3: PLAYER B chooses between RIGHT or LEFT, and the task ends.

Step 1: PLAYER A first chooses between IN or OUT.

- If PLAYER A chooses OUT, then the decision task ends. PLAYER A will receive \$6 and PLAYER B will receive \$2.
- If PLAYER A chooses IN, then the task proceeds to Step (2) where PLAYER A might privately learn the unknown information, and then have an opportunity to convey that information to PLAYER B. After Step (2), the task will proceed to Step (3) where PLAYER B will then be asked to decide between RIGHT or LEFT.

Step 2: I postpone the details about the information that PLAYER A can possible learn, and convey to PLAYER B until after Step (3) is described. Describing Step (3) first will help clarify Step (2).

Step 3: If PLAYER A chooses IN, at Step (1), then PLAYER B must choose between RIGHT or LEFT.

- If PLAYER B chooses RIGHT, then the decision task ends. PLAYER A will receive \$10 and PLAYER B will receive \$4.
- If PLAYER B chooses LEFT, then the decision task ends and PLAYER A will receive \$X and PLAYER B will receive \$6. There is a 50% chance that X = \$0 and a 50% chance that X = \$6. That is, X = \$0 and X = \$6 are equally likely.

NOTE: When the decision task begins, neither PLAYER A nor PLAYER B knows the value of X. Therefore, PLAYER A does not know the value of X when he/she decides between IN or OUT in Step (1). Now we proceed with the description of Step (2):

Step 2: If PLAYER A chooses IN, there is an 80% chance that PLAYER A will privately learn the value of X, and a 20% chance that PLAYER A will not privately learn the value of X.

- If PLAYER A does learn the value of X (80% chance), then PLAYER A must then decide whether or not to convey the value of X to PLAYER B before PLAYER B makes his/her decision in Step (3).

  - If PLAYER A does convey the value of X, then PLAYER B will know the value of X before he/she decides between RIGHT or LEFT in Step (3).
  - If PLAYER A does not convey the value of X, then PLAYER B will not know the value of X before he/she decides between RIGHT or LEFT in Step (3).
- If PLAYER A does not learn the value of X (20% chance), then PLAYER A will not have an opportunity to convey the value of X to PLAYER B. The task will proceed to step (3) where PLAYER B will then choose between RIGHT or LEFT without knowing the value of X.

Payoff Table:

The table below summarizes the earnings of each Player for each of the possible outcomes in the decision task:

Decision Outcome	PLAYER A Earnings	PLAYER B Earnings
PLAYER A chooses OUT	\$6	\$2
PLAYER A chooses IN and then:		
PLAYER B chooses RIGHT	\$10	\$4
PLAYER B chooses LEFT	\$$X$	\$6
There is a 50% chance $X = 0$ and a 50% chance $X = 6$		

Each person will participate in this decision making task ONE time. After the task has ended, the decision(s) of each Player and the corresponding earnings of each Player will be revealed to both Players. Additionally, the value of X will be revealed to both PLAYER A and PLAYER B regardless of the decisions made in the task. You will then be asked to fill out a short questionnaire that will take about 3 minutes to complete. Your answers to the questionnaire are confidential and will not be shared with any other participants. After completion of the questionnaire, an Experimenter will then come by and privately pay you your total experimental earning which equals your earnings from the decision task PLUS the \$5 show-up payment. After you have been paid, you may quietly exit the lab.

Sample Questionnaire - Player A

1. What is your Gender?
2. How old are you?
3. What is your major?
4. Have you ever taken an economics course?
5. Are you currently carrying more than $10 in cash?
6. Rate the amount of guilt you think Player B felt (would have felt) from choosing Left when (if) Player B knew the value of X was [true value of X]
7. Rate the amount of guilt you think Player B felt (would have felt) from choosing Left when (if) Player B did know the value of X.
8. How did you hear about the Economic Science Lab?
9. Have you ever referred a friend to the Economic Science Lab?
10. Is English your first language?

References

[1] Bacharach, M., Guerra, G., & Zizzo, D. (2007). "The Self-Fulfilling Property of Trust: An Experimental Study." Theory and Decision 63, 349-388.

[2] Battigalli, P., & Dufwenberg, M. (2007). "Guilt in Games." American Economic Review 97, 170-176.

[3] Battigalli, P., & Dufwenberg, M. (2009). "Dynamic Psychological Games." Journal of Economic Theory 144, 1-35.

[4] Baumeister, R., Stillwell, A., & Heatherton, T. (1994). "Guilt: An Interpersonal Approach." Psychological Bulletin 115, 243-267.

[5] Baumeister, R., Stillwell, A., & Heatherton, T. (1995). "Personal Narratives About Guilt: Role in Action Control and Interpersonal Relationships." Basic and Applied Social Psychology 17, 173-198.

[6] Beck, A., Kerschbamer, R., Qiu, J., & Sutter, M. (2010). "Guilt from Promise-Breaking and Trust in Markets for Expert Services – Theory and Experiment." University of Gothenburg Working Paper.

[7] Berg, J., Dickhaut, J. & McCabe, K. (1995). "Trust, Reciprocity and Social History." Games and Economic Behavior 10, 122-142.

[8] Bohnet, I., & Huck, S. (2004). "Repetition and Reputation: Implications for Trust and Trustworthiness when Institutions Change." American Economic Review Papers and Proceedings 94, 362-366.

[9] Bolton, G., Katok, E., & Ockenfels, A. (2005). "Cooperation Among Strangers with Limited Information About Reputation." Journal of Public Economics 89, 1457-1468.

[10] Bolton, G., & Ockenfels, A. (2000). "A Theory of Equity, Reciprocity, and Competition." American Economic Review 90, 166-193.

[11] Charness, G., Du, N., & Yang, C. (2011). "Trust and Trustworthiness Reputations in an Investment Game." Games and Economic Behavior 72, 361-375.

[12] Charness, G., & Dufwenberg, M. (2006). "Promises and Partnership." Econometrica 74, 1579–1601.

[13] Charness, G., & Dufwenberg, M. (2011). "Participation." American Economic Review 101, 1211-1237.

[14] Cox, J. (2004). "How to Identify Trust and Reciprocity." Games and Economic Behavior 46, 260-281.

[15] Duffy, J., Lee, Y., & Xie, H. (2008). "Social Norms, Information, and Trust Among Strangers: An Experimental Study." Mimeo.

[16] Dufwenberg, M. (2002). "Marital investments, time consistency, and emotions." Journal of Economic Behavior and Organization 48, 57-69.

[17] Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). "The Framing of Games and the Psychology of Play." Games and Economic Behavior (In Press).

[18] Dufwenberg, M. & Gneezy, U. (2000). "Measuring beliefs in an experimental lost wallet game." Games and Economic Behavior 30, 163-182.

[19] Dufwenberg, M. & Kirchsteiger, G. (2004). "A Theory of Sequential Reciprocity." Games and Economic Behavior 47, 268-298.

[20] Dulleck, U. & Kerschbamer, R. (2006). "On doctors, mechanics, and computer specialists: The economics of credence goods." Journal of Economic Literature 44, 5-42.

[21] Dulleck, U., Kerschbamer, R., & Sutter, M. (2009). "The economics of credence goods: On the role of liability, veri ability, reputation and competition." American Economic Review, forthcoming.

[22] Ellingsen, T., Johannesson, M., Tjotta, S., & Torsvik, G. (2010). "Testing Guilt Aversion." Games and Economic Behavior, 68, 95-107.

[23] Engelmann, D., & Fischbacher, U. (2009). "Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game." Games and Economic Behavior 67, 399-407.

[24] Fehr, E. (2009). "On the Economics and Biology of Trust." Journal of the European Economic Association 7, 235-266.

[25] Fehr, E., & Schmidt, K. (1999). "A Theory of Fairness, Competition, and Cooperation." Quarterly Journal of Economics 114, 817-868.

[26] Fischbacher, U. (2007). "z-Tree, Toolbox for Readymade Economic Experiments." Experimental Economics 10, 171–178.

[27] Fong, Y., Huang, C., & Offerman, T. (2007). "Guilt Driven Reciprocity in a Psychological Signaling Game." Working Paper.

[28] Geanakopolos, J., Pearce, D. & Stacchetti, E. (1989). "Psychological games and sequential rationality." Games and Economic Behavior 1, 60-79.

[29] Greiner, B., & Levati, V. (2005). "Indirect Reciprocity in Cyclical Networks: An Experimental Study." Journal of Economic Psychology 26, 711-731.

[30] Lee, S.Y., (2010). "Economics of Guanxi as an Interpersonal Investment Game." Review of Development Economics 14, 333-342.

[31] Rabin, M. (1993). "Incorporating fairness into game theory and economics." American Economic Review 83, 1281-1302.

[32] Reuben, E., Sapienza, P., & Zingales, L. (2009). "Is Mistrust Self-Fulfilling?" Economic Letters 104, 89-91.

[33] Seinen, I., & Schram, A. (2006). "Social Status and Group Norms: Indirect Reciprocity in a Helping Experiment." European Economic Review 50, 581-602.

[34] Tangney, J., & Fischer, K. (1995). Self-conscious emotions: Shame, guilt, embarrassment, and pride. New York: Guilford Press.

[35] Vangelisti, A., Daly, J., & Rudnick, J. (1991). "Making People Feel Guilty in Conversations: Techniques and Correlates." Human Communication Research 18, 3-39.