Price and Quality with a Conscientious Worker

Ben Zamzow

November 5, 2012

Preliminary draft; please see http://u.arizona.edu/~zamzow for current version.

In this paper I introduce guilt aversion into an effort-setting game in a way not yet considered by labor economists and make interesting predictions. I also raise a word of caution for the application of Battigalli & Dufwenberg (2007, AER) simple guilt preferences and show that for my game an adaptation is necessary. I begin with a sequential game played between a firm, a consumer, and a worker. Under the assumption that all players are selfish and rational, I obtain the baseline prediction of a unique sequential equilibrium that is inefficient. I demonstrate using standard methods that the firm can enact a program of monitoring and penalization in order to incentivize high effort from the worker thereby supporting the efficient equilibrium. Alternatively the firm may rely on guilt aversion to motivate the worker as introduced by BD (2007). With sequential equilibrium and extensive form rationalizability I show that the predictions of the resulting model are inconsistent with reasonable intuition. I propose an adaption of BD (2007) simple guilt for application in the context of this game and demonstrate that extensive form rationalizability is the appropriate solution concept fully compatible with economic intuition. Finally, I observe that in order to achieve the efficient outcome, the profit-maximizing firm prefers inducing guilt in the worker to the alternative of using monitoring and penalization.

1 Introduction

The temptation for workers to shirk can have great negative consequences for the functioning of markets especially wherein the quality of the good is determined by a worker’s effort exertion. Here certainly the expected effort level of the worker influences the strategic behavior of both the firm and consumer. Since consumers may refuse to buy in the absence of
trust that the good will be of high enough quality, such a firm unable to guarantee sufficient quality quite likely will not stay in the market. In the absence of additional monitoring and oversight by the firm it seems that market failure is inevitable, however, it is reasonable that guilt aversion can change the predictions.\(^1\) In particular I am interested in the influence of guilt aversion on effort choices of a worker that then affects consumer and firm behavior. Rather than examine the case where a worker wishes not to disappoint his employer, I consider the case where he wishes not to disappoint the consumer. I am therefore interested in how the worker’s effort choice under guilt aversion affects the behavior of the firm and consumer.

One context where this form of guilt aversion enters is a market for a made-to-order good. Examples of this are things like concession stands, home repairs, and shoeshines. In general these markets work as follows. A firm employs a worker who is asked to be available to create a good upon consumer request. The worker typically receives his wage regardless of the purchase decision made by the consumer. If the consumer requests a product, the worker must exert effort to create it. Crucially, the quality of the final product received by the consumer depends on the effort exertion of the worker. The following vignette will help place the context within which I am working and also introduces the relevant economic and psychological intuition:

Consider Bill, a worker who is paid a wage to operate a concession stand at municipally owned Beach Bay Amusement Park. No other firm or vendor is permitted to operate on the grounds and so the amusement park has the sole ability to sell to its attendees. His responsibilities include making and selling cotton candy and snow cones. When customers come to his window he takes their order and accepts payment before making the desired treat. Both snacks require preparation that is mildly unpleasant and Bill’s chosen effort level determines the quality of the treat the customer actually receives. The cotton candy machine is loud and hot causing wisps of sugar to fly in his eyes and hair. He must take care to make sure the sugar mass adheres to the paper stick and has an appropriate level of fluffiness. The shaved ice from the snow cone machine freezes his fingers and he must keep it well stocked with fresh ice lest the product be watery. He must take care to make sure the flavoring is evenly dispersed and fills the bottom of the cone or else the treat may have a pleasant appearance but will taste bland. Of course there is a bare minimum effort level and associated quality level in order that the product actually constitutes cotton candy or snow

\(^1\)Other authors have offered models with predictions that avoid shirking. A well known example is the Shapiro-Stiglitz (1984) model in which firms pay an efficiency wage which leads to a no shirking equilibrium. This prediction arises in part because high wages create unemployment which makes the prospect of firing very unattractive to the worker. More recently Dufwenberg & Kirchsteiger (2000) show that reciprocally behaving workers will interpret high wages as kind and therefore respond with high effort.
cone and below which he would certainly be fired. However, Bill believes that choosing the minimum effort level means also that the resulting cotton candy or snow cone will disappoint his young customer. To the extent that Bill is sensitive to guilt he sets a high effort level so that he believes his creation meets the expectations of the consumer.\footnote{Imagine a young customer comes to Bill’s window and orders a cotton candy. He may notice that the child is paying with crumpled bills and coins that had likely been squirreled away in a piggy bank for a special occasion. He believes this customer has high hopes for a very nice cotton candy. Under such circumstances it is reasonable for Bill to imagine that a small, matted cotton candy is quite disappointing. Further, Bill realizes that when he delivers the treat he will either see an expression of happiness or a downcast look.}

To capture this intuition I introduce a simple three player game and compare predictions under three policies. I show that when all players are selfish there is a unique sequential equilibrium that is inefficient. I demonstrate that if the firm incentivizes higher effort through the use of a monitoring and penalty system the efficient equilibrium arises. Finally, I show that this efficient equilibrium might be available at a lower cost to the firm when the worker is guilt averse. To this end, I work within the framework of psychological game theory and allow the worker to have belief dependent preferences. In the process I call attention to an issue pertaining to the direct application of the Battigalli & Dufwenberg (2007) model of simple guilt and I suggest an adaption for my game. Finally I compare the predictions under guilt aversion to those with the monitoring scheme.

2 Background

I model the basic interaction as a simple three player game that begins with a move by the firm, either \textit{In} or \textit{Out}. If the firm remains in the market by choosing \textit{In} the consumer is called upon to move. On the other hand if the firm exits by choosing \textit{Out} the game ends immediately. The consumer observes the choice of the firm and selects \textit{Buy} or \textit{¬Buy}. If the consumer (she) declines to make a purchase the game ends and otherwise the worker (he) receives the move. The worker observes the decisions of both previous players and then decides to exert high or low effort by choosing $e_h$ or $e_l$ respectively. I assume the selection of $e_h$ also implies a better product is created.

Payoffs are assigned according to the extensive game displayed in the left panel of Figure 1. If the game ends with the firm exiting the market the firm and worker are assigned payoffs of zero while the consumer receives $x$. The parameter $x$ corresponds to the value the consumer receives in the absence of the opportunity to buy from the firm. Although it stands to reason this should have no influence on the analysis I will vary this parameter to make important observations later on. If play proceeds to her decision node and the consumer chooses not to purchase the good, she receives a payoff of zero. I assume the firm makes a
wage payment, \( w \), regardless of the consumer’s purchase decision and therefore the payoffs to the firm and worker at this end node are \( -w \) and \( w \) respectively. Finally at the two end nodes following the worker’s effort choice, the payoff to the firm is equal to the price of the good minus the wage paid to the worker. The payoff to the consumer is the value of the good minus its price, \( v_i - p \) where the worker’s selection of \( e_i \) implies a value of \( v_i \) for \( i \in \{ l, h \} \). Finally, the payoff to the worker is his wage minus the cost of his effort level. I represent the costs associated with high and low effort selections with \( c_h \) and \( c_l \) respectively.

For the remainder of the paper I will focus on specific numerical examples in order to streamline the presentation. The first one I will consider is \( \Gamma_1 \) which arises when the parameters are assigned the following values: \( w=2, p=5, v_h=7, v_l=4, c_l=0, c_h=1 \). This game is displayed in the right panel of Figure 1.

![Figure 1](image)

### 2.1 Classical Analysis

As a baseline reference it is useful to determine the prediction when all players are selfish and rational. The strategy profile \((\text{Out}, \neg \text{Buy}, e_l)\) is the backward induction solution and also the unique sequential equilibrium (SE) for all \( x \). This equilibrium is inefficient since relative to the selfish prediction all players are better off with \((\text{In}, \text{Buy}, e_h)\) when \( x \leq 2 \). The latter would occur as an equilibrium if players were to reasonably expect the worker to select \( e_h \) upon receiving the move and if he indeed selects high effort when called upon. Clearly this does not happen in the present model, but it may if additional structure is added.
2.2 Monitoring and Penalty Scheme

In this section I alter $\Gamma_1$ to allow the firm to exert some control over the worker. In $\Gamma_1$ if the firm were somehow able to ensure a high effort selection from the worker, the consumer would surely buy and the firm would receive a positive payoff. One way in which this is possible is if the firm were to monitor the worker and penalize low effort to the extent that the worker always exerts high effort. To bring this alive in the model suppose the firm is able to monitor the worker at cost $\mu$ and penalize low effort by $\tau$. The resulting game, $\Gamma_2$, is shown in Figure 2. Certainly in this new game as long as $\tau > 1$ the worker prefers to play $e_h$. Moreover, when the monitoring cost is such that $\mu < 3$, the strategy profile $(In, Buy, e_h)$ is a SE for appropriate beliefs.

Figure 2. $\Gamma_2$

While this structure makes the nice equilibrium available, there is reason to be concerned that monitoring may not be optimal for the firm to pursue in this context.\(^3\) It may be possible for the firm to achieve the efficient outcome instead by relying on the conscience of the worker. If so, this could quite likely be accomplished at a lower cost to the firm than any sort of monitoring and penalty system. In fact, I will show that with a sufficiently guilt averse worker the firm no longer needs to monitor.

\(^3\)There is a large literature on the effects of monitoring in principal-agent games. Frey (1993) observes that monitoring may have either a disciplining effect that increases effort or a crowding-out effect that decreases effort depending on the nature of the relationship between the firm and worker. Falk & Kosfeld (2006) find experimental evidence suggesting that agent effort selections decrease with increased monitoring in the sense of setting a minimum effort standard. Dickinson & Villeval (2008) present an experimental test of Frey (1993) in a principal-agent game and find that in the context of an interpersonal relationship monitoring has a negative effect on effort when the principal’s payoff is increasing in the worker’s effort.
3 Guilt Aversion

To this point in the paper I have assumed that all players have selfish preferences. Henceforth, I will endow the worker with guilt aversion in the subgame that includes his decision node. This will have the following interpretation. When the worker believes the consumer receives a product of lower value than he believes was initially expected the worker is affected by guilt. To the extent the worker anticipates feeling guilty for the consequence of his shirking he may instead be motivated to select the higher effort choice.

In order to capture the correct intuition it is helpful to return to the vignette. One can imagine a little girl coming to Bill’s window and ordering a cotton candy. He believes the little kid is expecting a stupendously fluffy, full cotton candy and will be sad upon receiving a small, matted one instead. While Bill finds high effort distasteful he also dislikes believing he has let down the consumer. Furthermore, Bill knows he will see the expression on the child’s face as he delivers the cotton candy to her. In order to avoid feelings of guilt, Bill decides instead to invest higher effort so that the end result matches his belief about what the child expects.

At this point it is important to pause and emphasize that I will be assuming that the worker has belief dependent utility. For this reason the use of psychological game theory is warranted in order to analyze this preference structure in a systematic framework. In this way I will be able to separate the worker’s payoffs into material and belief-dependent, psychological components. The material component will just be the monetary payoffs shown in the game tree. In order to construct the psychological component of the payoff I will apply the model of simple guilt preferences from Battigalli & Dufwenberg (2007).

3.1 Incorporating Guilt Aversion

In this section I will introduce the mathematical structure useful to incorporate guilt aversion into the preferences of the worker. The formal treatment I begin with can be found in Battigalli & Dufwenberg (2007), henceforth BD (2007), which presents a general model of preferences incorporating guilt aversion and a solution concept that adapts sequential equilibrium (SE) to psychological extensive games.\(^4\) The proper mathematical structure within which my paper fits is Battigalli & Dufwenberg (2009), henceforth BD (2009), which presents a general framework for psychological games and, quite importantly for my purposes, adapts

---

\(^4\)Guilt preferences were introduced into psychological game theory in Dufwenberg (2002) which applies psychological forward induction reasoning to get predictions in a simple two player extensive form game. This paper draws inspiration from Dufwenberg (2002) but instead considers a game with three players and, in addition, models guilt slightly differently.
extensive form rationalizability for application in psychological games.\textsuperscript{5} To help familiarize the reader I will provide a brief description of BD (2007) simple guilt preferences, however, the following will provide the minimal explanation necessary and I refer the interested reader to the original for the full formalization.

In my paper the worker is affected by guilt aversion in his subgame and nowhere else. The other two players will retain selfish preferences throughout. For these reasons my treatment of BD (2007) simple guilt preferences is presented with respect to the worker only.\textsuperscript{6} The worker lets down or disappoints the consumer if as a result of his strategy choice the consumer gets a lower monetary payoff than she had initially expected.

To operationalize this in my game, the consumer has a probability distribution over strategies and therefore has an initial conditional expected payoff, call $E_{o,c}[\hat{\pi}_c | h^o]$ where the initial history, or the root, is denoted by $h^o$. The worker does not have access to $E_{o,c}[\hat{\pi}_c | h^o]$ since this is a belief held by the consumer. Instead the worker forms his own initial expectation of the consumer’s expected payoff which I will denote $E_{wc}^{o} [\hat{\pi}_c | h^o]$. The worker’s initial expectation of the consumer’s belief is therefore the amount of payoff the worker initially believes the consumer initially expects to receive. At his own decision node the amount by which the worker believes he disappoints the consumer is given by the difference $\max\{0, E_{wc}^{o} [\hat{\pi}_c | h^o] - \pi_c\}$ where $\pi_c$ is the actual payoff received by the consumer. That is to say that the consumer is disappointed when she receives a payoff lower than what she had initially expected to receive. The worker also has a sensitivity to guilt, $\theta_{wc}$, governing the degree to which disappointment affects him. It is assumed that $\theta_{wc} \geq 0$ with equality meaning that guilt has no bearing on behavior. The utility function of the worker with simple guilt preferences is given by:

$$u_w = \pi_w - \theta_{wc} \cdot \max\{0, E_{wc}^{o} [\hat{\pi}_c | h^o] - \pi_c\}$$

The game that arises from $\Gamma_1$ when the worker is guilt averse is displayed in Figure 3. I refer to this new game as $\Gamma_3(x)$ where $x$ denotes the parameter in the consumer’s payoff following the firm’s decision to exit the market. In the game tree, the term $\psi_{wc}$ represents the psychological component of the worker’s payoff as discussed above. Observe that the worker’s payoff at the end node following $e_l$ is not a number but actually a function of beliefs.

\textsuperscript{5}Recent work in the area of psychological game theory and guilt preferences has not been confined to theory. Several papers have investigated the predictions of guilt aversion in the laboratory. An early example is Dufwenberg & Gneezy (2000); more recently there have been papers testing the predictions of BD (2007) guilt aversion in the laboratory including Charness & Dufwenberg (2011) and Cardella (2012).

\textsuperscript{6}BD (2007) actually present two versions of guilt aversion, simple guilt and guilt from blame. In my paper I will apply the definition of simple guilt which holds that the worker’s guilt depends on how much he believes he disappoints the consumer. On the other hand, guilt from blame holds that the worker’s guilt also depends on how much he believes the consumer believes he intended to let her down.
The payoff to the worker at this end node is crucially belief-dependent. For this reason the new game, $\Gamma_3(x)$, does not belong to the class of games considered by standard game theory. Instead, $\Gamma_3(x)$ is a psychological game of the variety introduced by Geanakoplos, Pearce, & Stacchetti (1989) and further studied by Battigalli & Dufwenberg (2009).

### 3.2 Insights from Battigalli & Dufwenberg (2009)

While the natural way to bring in guilt aversion is to follow BD (2007), in this game there will actually be a difficulty that arises from the direct application of these simple guilt preferences. In order to fully appreciate the issue I will motivate the problem by transporting the logic from an example found in BD (2009) into two variations of $\Gamma_3(x)$. This game has similar properties to $\Gamma_3(x)$ and will provide some insight as to what might happen in my game with the addition of simple guilt. I will summarize their results and I direct the interested reader to the original for additional details. The BD (2009) trust game is displayed in Figure 4. There are two players in the game, Ann and Bob. Ann moves first and chooses either to continue the game by selecting Trust or to end the game by choosing Don’t. Payoffs following her choice of Don’t are $(1, 1)$. If she chooses to Trust it becomes Bob’s move and he may Share or Grab. Payoffs following Share are $(2, 2)$ and payoffs following Grab are $(0, 4 - 5\beta)$. The term $5\beta$ appearing in Bob’s payoff following Grab reflects guilt he feels for having disappointed his co-player. Some additional details are necessary to make clear exactly how this works. Ann holds a beliefs $\alpha$ that Bob will share if she trusts, however, Bob does not have access to this belief. Instead Bob knows $\beta$ which is his own expectation of Ann’s belief, $\alpha$. The 5 captures Bob’s sensitivity to guilt.
I begin the analysis by searching for SE. At first it may seem uncontroversial that when $\beta$ is quite large and when Bob dislikes disappointing Ann one should expect Bob to Share whenever Ann Trusts. This actually turns out not necessarily to be the case. Certainly when Bob’s guilt aversion is high and both players initially expect to follow the strategy profile (Trust, Share), Bob will not deviate when he is called upon to play. He believes Ann initially expects a payoff of 2 and when he is highly guilt averse will not let her down. However, BD (2009) explain that in equilibrium players will never update their beliefs about the beliefs held by co-players. Therefore with SE reasoning if Bob receives the move unexpectedly he is incapable of holding beliefs that Ann expects him to Share. Bob is able to update his belief about the strategy that Ann is playing however he does not update about beliefs. For this reason if Bob initially expects that Ann will select Don't and he nevertheless finds himself at his decision node, he will select Grab regardless of his sensitivity to guilt. This being the case, for high sensitivities to guilt, there are multiple SE in the BD (2009) trust game. In light of this observation SE is actually somewhat unattractive when applied in psychological games and it may be more appropriate to employ a solution concept that allows for psychological forward induction reasoning.\footnote{Dufwenberg 2002 introduces psychological forward induction to model the effects of guilt in a simple marriage-investment game.}

BD (2009) also determine the predictions of extensive form rationalizability which nicely captures the economic and psychological intuition that one may hope to model in this game. Contrary to the ambiguous prediction with sequential equilibrium they find a unique rationalizable outcome when Bob has a high sensitivity to guilt. Before explaining this result I will explain extensive form rationalizability as discussed in BD (2009). The key is the insistence on strong belief. If a proposition is strongly believed, players must always revise beliefs so as to maintain belief in that proposition as long as feasible. Relevant here is the insistence on strong belief in rationality. Strong belief in rationality has the consequence...
that whenever Bob is called upon to move he must believe Ann expects him to Share in order to retain his belief that she is rational. Whenever he receives the move, Bob believes Ann initially expects to receive a payoff of 2. Therefore when Bob’s guilt aversion is high, he will in fact Share, and so extensive form rationalizability returns a unique prediction in the BD (2009) trust game.

To summarize the findings for high guilt sensitivity, when Bob has a strong sensitivity to guilt, the prediction of sequential equilibrium is ambiguous as either outcome is possible while the prediction of extensive form rationalizability is unique, Bob will always Share.

3.3 Example of $\Gamma_3(0)$ with guilt averse worker

Having gained insights from the analysis of the BD (2009) trust game I will now turn to $\Gamma_3(x)$. I will consider two versions corresponding to when parameter $x$ takes on values of 0 and $-1$. First I analyze $\Gamma_3(0)$. Intuitively the reader may recognize that the value of $x$ should actually have no bearing on the behavior of the players in the game. It should be irrelevant to the consumer’s expected payoff following the decision to purchase and therefore should not affect the degree to which the worker believes his actions potentially disappoint the consumer. However, it will turn out that when the worker is endowed with BD (2007) simple guilt preferences, which insists on referring to initial beliefs for assessments of guilt, the predictions are actually not invariant to changes in the value of $x$. I will discuss this in greater detail later.

In light of the findings with BD (2009)’s trust game it is reasonable to expect multiple sequential equilibrium and a unique extensive form rationalizable outcome. But actually, the situation turns out to be much more complex. By calling attention to the predictions of both solution concepts for large guilt sensitivities, $\theta_{wc}$, I will demonstrate that some care is warranted when applying BD (2007) simple guilt preferences. An immediate observation is that when $\theta_{wc}$ is low guilt aversion has no effect and a unique equilibrium is predicted matching the classical prediction with both SE and extensive form rationalizability. For this reason my treatment will henceforth ignore the case of low $\theta_{wc}$.

What happens when $\theta_{wc}$ is high enough to influence behavior? The strategy profile, $[In, Buy, e_h]$, can be supported as a SE for $\theta_{wc} \geq \frac{1}{3}$. To see this note that no player will deviate. At their decision nodes the firm and consumer earn positive payoffs by remaining in the game and zero otherwise. More explanation is necessary for the worker. Observe that upon receiving the move the worker consults his own beliefs about the consumer’s initial expected payoff which is 2. If the worker selects low effort the consumer will receive a payoff of $-1$, if he selects high effort her payoff will indeed be 2. The worker then selects high effort

10
when $2 - \theta_{wc}(3) \geq 1$ which is true for $\theta_{wc} \geq \frac{1}{3}$.

On the other hand, can an equilibrium be sustained where the firm goes Out? It depends on sensitivity to guilt. The strategy profile $[Out, \neg Buy, e_t]$ means that the consumer expects payoff $x = 0$. In order for this to be an equilibrium it must be the case that the worker chooses $e_t$ if called upon. What actually happens off the equilibrium path? When he unexpectedly receives the move the worker is able to update his beliefs about the consumer’s strategy but not about her beliefs. This follows as a consequence of the logic of sequential equilibrium. BD (2009) explain that in equilibrium players will never update their beliefs about the beliefs held by co-players. Therefore, when assessing the degree to which he believes the consumer could be let down the worker checks his initial beliefs about the consumer’s initial expected payoff which is $x = 0$. The low effort selection is made when $2 - \theta_{wc}(1) < 1$ or when $\theta_{wc} < 1$.\(^8\) When $\theta_{wc} \geq 1$ there is a unique sequential equilibrium corresponding to the strategy profile, $[In, Buy, e_h]$.

These findings run contrary to what might be expected in two regards. First, the prediction of sequential equilibrium for high guilt sensitivities differs from the analogous prediction in the BD (2009) trust game in which multiple sequential equilibrium were predicted. Second, the fact that this result fails to obtain happens due to a feature of the three player structure of the game and in part due to an artifact of the consumer’s payoffs. After all, when he is surprised with the move the logic of sequential equilibrium prevents the worker from believing that the consumer expects him to exert high effort and therefore should protect him from being vulnerable to guilt. But, for large enough sensitivity to guilt he chooses $e_h$ anyway. This is because as a consequence of BD (2007), the worker must live up to his initial belief about the consumer’s initial expected payoff which here is $x = 0$. But this is not the correct psychology. Bill cares about what the child expects when she purchases the cotton candy, not her initial expectations, $x$, at some time prior to the amusement park’s decision to open! In the next section I will discuss how predictions differ when the parameter $x$, and therefore initial beliefs, take on a different value. First, however, I will discuss the predictions of extensive form rationalizability.

Recall extensive form rationalizability imposes the requirement of strong belief in rationality. Here this means that even when receiving the move unexpectedly the worker must revise his beliefs so as to maintain his belief that the consumer is rational. The worker actually gets the rationality correct but this has no bearing on his own best response because he cares about the initial belief. Unlike in BD (2009) strong belief in rationality does not force the worker to change belief of payoffs. But simple guilt cares about the initial beliefs. Further it happens that the parameter $x$ shapes predictions here.

\(^8\)The prediction of SE is ambiguous when $\theta_{wc}$ takes on an intermediate value, $\theta_{wc} \in [\frac{1}{3}, 1)$. 

11
The strategy profile $[In, Buy, e_h]$ is extensive form rationalizable for $\theta_{wc} \geq \frac{1}{3}$. To see this observe that a rational firm and rational consumer will both remain in the game. At his decision node the consumer is able to signal the expectation of at most 2 when selecting $Buy$. Upon receiving the move the worker consults his belief about the consumer’s expectation. He believes a rational consumer expects a payoff of 2 and therefore he chooses between $e_l$ and $e_h$ according to: $2 - \theta_{wc}(3) \geq 1$ which is true for $\theta_{wc} \geq \frac{1}{3}$.

Can an equilibrium be sustained in which the firm goes $Out$? This requires that the worker selects $e_l$ if called upon to move. For this to occur it must be the case either that the worker believes the consumer expects him to exert low effort, or the worker believes the consumer expects high effort exertion, but is insufficiently motivated by guilt aversion to care. Taken together the strategy profile $[Out, \neg Buy, e_l]$ is extensive form rationalizable for $\theta_{wc} < 1$. To see this observe that when the worker receives the move unexpectedly in order maintain his belief in the consumer’s rationality he believes that she expects a payoff of at least 0, the payoff received by the her choice of $\neg Buy$. However, while the worker correctly rationalizes the consumer’s choice, for the purposes of assessing potential disappointment the worker looks to the consumer’s initial expectation, $x = 0$. It is this belief that the worker must live up to. The worker gives the consumer a payoff equal to her initial expectation when $e_h$ is chosen with probability one-third. The worker chooses $e_l$ when $2 - \theta_{wc} > 1$ or when $\theta_{wc} < 1$. For high guilt sensitivity, $\theta_{wc} \geq 1$, there is a unique prediction of $[In, Buy, e_h]$.

The argument provided in the foregoing discussion demonstrates that the model actually portrays the wrong psychology. While the worker correctly rationalizes the consumer’s decision to $Buy$, again it is the initial belief that is driving behavior. But, rather than care about the initial belief, it stands to reason the worker should care about the conditional belief which is signaled by the consumer’s decision to $Buy$. Since it is the initial belief that matters here, if $x$ took on other values the prediction would change! Therefore in some sense extensive form rationalizability delivers the correct prediction, but based on the wrong reasons.

In summary, the results of the analysis of the game with both sequential equilibrium and extensive form rationalizability are partially contrary to what one may expect in light of the findings in the BD (2009) trust game. Unlike with the trust game I instead showed that a unique prediction arises from SE for high $\theta_{wc}$. On the other hand, similar to the trust game a unique prediction arises from extensive form rationalizability for high $\theta_{wc}$. While a unique prediction was obtained with SE for high $\theta_{wc}$ things could have been different for other values of $x$. I will now demonstrate this with a concrete example by considering a slightly different game created by setting $x = -1$.

---

9For values of $\theta_{wc} \in \left[\frac{1}{3}, 1\right)$ the predictions of extensive form rationalizability are ambiguous.
3.4 Example of $\Gamma_3(-1)$ with guilt averse worker

I now consider $\Gamma_3(-1)$ which arises when $x = -1$. What happens for high values of $\theta_{wc}$? The strategy profile, $[In, Buy, e_H]$ can be supported as SE for $\theta_{wc} \geq \frac{1}{3}$. To see this observe that when it becomes his move the worker must live up to his belief about the consumer’s initial expected payoff of 2. The consumer receives a payoff of $-1$ when he chooses $e_l$ and she receives a payoff of 2 when he chooses $e_h$. The worker exerts high effort when $2 - \theta_{wc}(3) \geq 1$ or when $\theta_{wc} \geq \frac{1}{3}$.

But, the strategy profile $[Out, \neg Buy, e_l]$ can be supported as a SE for $\theta_{wc} \geq \frac{1}{3}$ as well. When the worker unexpectedly finds himself with the move he is able to update his beliefs about the strategies the consumer played but not about her beliefs held at her decision node. Since BD (2007) simple guilt uses initial beliefs the worker must live up to his belief about the initial expected payoff of $-1$. However the consumer gets $-1$ when the worker chooses $e_l$ which means that guilt aversion presents no reason to exert high effort. Therefore the worker selects $e_l$ if $2 - \theta_{wc}(0) \geq 1$, which is always true! This shows that for values of $\theta_{wc}$ above one-third the prediction of sequential equilibrium is ambiguous.

What about the predictions of extensive form rationalizability? The strategy profile $[In, Buy, e_h]$ is extensive form rationalizable for $\theta_{wc} \geq \frac{1}{3}$. To see this observe that when players maintain belief in co-player rationality all players are best responding. For the firm $3 > 0$, for the consumer $2 > 0$, and for the worker $1 - \theta_{wc}(3) \geq 2$ assuming $\theta_{wc} \geq \frac{1}{3}$.

However, the strategy profile $[Out, \neg Buy, e_l]$ is also extensive form rationalizable for high values of $\theta_{wc}$. This either requires that the worker believes the consumer expects a payoff of $-1$ or that the worker believes the consumer expects a payoff of 2, but is insufficiently motivated by guilt aversion for it to influence his behavior. The worker is always able to rationalize the belief that the consumer expects him to select $e_l$. To see this suppose that the worker finds himself with the move unexpectedly. Consulting his initial beliefs the worker believes the consumer has initially expected $x = -1$. Since the consumer gets $-1$ when the worker chooses $e_l$ and 2 when he chooses $e_h$ the worker can maintain his belief in the consumer’s rationality and continue to believe that she expects him to select low effort. For these reasons despite potentially high values of $\theta_{wc}$ the worker is able to choose $e_l$ without guilt aversion having an effect. Therefore this shows that extensive form rationalizability yields an ambiguous prediction for $\theta_{wc} \geq \frac{1}{3}$.

The preceding demonstrates that when the behavior of the worker is driven by initial rather than conditional beliefs, the predictions of the model depend on the value of the parameter $x$. When $x = 0$, the guilt averse worker will select $e_h$ whenever called upon, however, when $x = -1$ the guilt averse worker is unable to be affected by guilt and is therefore may choose $e_l$. Either way the model captures incorrect intuition resting on initial
beliefs which should plausibly be immaterial to the interaction in the subgame between the consumer and worker.

In summary both sequential equilibrium and extensive form rationalizability yielded ambiguous predictions for high values of $\theta_{wc}$. Contrast this with the analysis of the trust game of BD (2009) wherein the SE prediction was ambiguous and rationalizability yielded a unique outcome. Moreover, these results differ from the predictions of $\Gamma_3(0)$ where a unique prediction arises for both solution concepts. The only difference was the value of $x$. So why are the predictions not invariant to changes in the parameter $x$? In the trust game Ann’s initial beliefs and those she holds at the time of her move coincide. But in $\Gamma_3(x)$ due to the presence of the firm the initial beliefs of the consumer and those held at the time of her move are decoupled. Further, the payoffs to the consumer at the end node prior to her decision affect the worker’s guilt assessment. The reason why this matters is the reliance on initial beliefs for simple guilt. The worker’s belief about the consumer’s initial expectation at the root determines extent of disappointment. But this seems a bit unreasonable. Not only should the prediction be invariant to changes in $x$ but more precisely it stands to reason that in this game it might be more appropriate to allow the worker to assess his guilt relative to the beliefs the consumer holds at the time of his move. I make a natural adaptation to the model by allowing for this in $\Gamma_3(x)$. In the next section I will implement this adaptation and discuss the predictions.

4 Adaptation to BD (2007) Simple Guilt

The model of simple guilt developed in BD (2007) can be modified as follows. Whereas previously the definition of simple guilt relied on the worker’s beliefs about what the consumer initially expects now I make the relevant belief that which the consumer holds at the time of her move. At the time of her move the consumer has a probability distribution over strategies and therefore she has a conditional expected payoff given her strategy and beliefs. This is denoted by: $E_c[\hat{\pi}_c|h]$. The worker does not have access to this belief since it belongs to the consumer. Instead the worker forms his own expectation of this belief which I denote $E_{wc}[\hat{\pi}_c|h]$. I then write:

$$u'_w = \pi_w - \theta_{wc} \cdot \max\{0, E_{wc}[\hat{\pi}_c|h] - \pi_c\}$$

As before the firm and consumer are assumed to retain selfish preferences and guilt aversion will become relevant in the subgame corresponding to the worker’s effort decision.
4.1 Results with simple guilt adapted for $\Gamma_3(x)$

In this section I will apply the adapted version of simple guilt to the $\Gamma_3(x)$. Now that this adjustment has been made it remains to go back and determine whether this makes a difference. A crucial observation is that the predictions of the model should be invariant to changes in the parameter $x$. With the adaptation to BD (2007) preferences the parameter $x$ actually will be irrelevant. I therefore no longer consider $\Gamma_3(0)$ and $\Gamma_3(-1)$ separately.

So what are the predictions of SE in $\Gamma_3$ when the worker has the adapted version of simple guilt preferences? As in the BD (2009) trust game there are multiple equilibria for high guilt aversion. This result follows as a consequence of the no belief updating feature of SE and is shown below in Claim 1.

**Claim 1** In $\Gamma_3$ for $x \in \{-1, 0\}$, and $\theta_{wc}$ is high, when the worker is endowed with BD (2007) simple guilt preferences adapted to depend on beliefs held at the time of the consumer’s move:

- (i) When $\theta_{wc} \geq \frac{1}{3}$ there are multiple SE; (a) $(In, Buy, e_h)$ and (b) $(Out, \neg Buy, e_i)$ are possible.
- (ii) When $\theta_{wc} < \frac{1}{3}$ the SE is $(Out, \neg Buy, e_i)$.

**Proof.**

- To verify (ia) it is enough to be sure there are no profitable deviations for any player. Receiving the move the worker believes the consumer expects a payoff of 2 at the time of her move. Given consistent beliefs the worker then calculates $2 - \theta_{wc}(3) \leq 1$ and so the worker indeed chooses $e_h$ when $\theta_{wc} \geq \frac{1}{3}$. To verify (ib) suppose at the root play is expected to follow $[Out, \neg Buy, e_i]$. The worker believes that at the time of her move the consumer expected a payoff of 0 which is attainable when the consumer uses $\neg Buy$. However, the consumer has nevertheless instead chosen $Buy$. Finding himself with the move unexpectedly the worker consults his beliefs about the consumer’s expectations. The worker believes that the consumer still expects him to choose $e_i$; he is incapable of holding beliefs making him susceptible to guilt. The worker chooses $e_i$ regardless of $\theta_{wc}$.

- To verify (ii) show that there are no profitable deviations for any player. When guilt aversion is weak the classical solution obtains. The worker never selects $e_h$ so the consumer never $Buy$ and the firm never goes $In$.

The above demonstrates that as with the trust game in BD (2009), the analysis of $\Gamma_3$ using sequential equilibrium yields ambiguous predictions for high values of $\theta_{wc}$. The
psychology of guilt aversion matches the vignette, but the logic of sequential equilibrium leads to predictions that do not match the intuition. With SE when Bill finds himself with the move unexpectedly he is not able to hold the belief that the worker expects high effort and therefore even with high guilt sensitivity, Bill is able to exert low effort. Recall that with ordinary simple guilt preferences this did not happen, but another issue arose. The prediction there was of a unique sequential equilibrium in which the worker chose high effort in order to live up to an initial belief.

So what happens when extensive form rationalizability is applied to $\Gamma_3$? With the adaptation to BD (2007) simple guilt preferences a forward induction argument goes through. The consumer is able to signal an expected payoff the belief about which the guilt averse worker must live up to. The prediction is for a unique equilibrium corresponding to the strategy profile $[In, Buy, e_h]$ when the worker is highly guilt averse.

Claim 2 $In \Gamma_3(x)$ for $x \in \{-1, 0\}$ when the worker is endowed with BD (2007) simple guilt preferences adapted to depend on beliefs held at the time of the consumer’s move and $\theta_{wc}$ is high:

- (iii) There is a unique psychological extensive form rationalizable equilibrium, $[In, Buy, e_h]$
- (iv) For low values of $\theta_{wc}$ the prediction is $[Out, \neg Buy, e_l]$.

Proof. Suppose play is expected to follow $(In, Buy, e_h)$. When the consumer chooses $Buy$ she signals the belief in a payoff of 2. The worker believes that the consumer believes he will choose $e_h$ with probability 1. Given the adapted version of simple guilt preferences, the worker pays attention to the consumer’s expectation held at the time of her move when assessing guilt, an anticipated payoff of 2. The consumer receives a payoff that is below this expected value when low effort is chosen and one that exceeds this expected value when high effort is selected. Therefore the worker selects high effort when $2 - \theta_{wc}(3) \leq 1$, or when $\theta_{wc} \geq \frac{1}{3}$. Given common strong belief in rationality the consumer (and firm) must also figure this out.

Suppose initially each player believes play will follow $(Out, \neg Buy, e_l)$, but the worker is unexpectedly called upon to play. Observe when the firm selects a price offer it signals to the consumer an expectation of a payoff of at least 0 which could have been obtained by going out. Maintaining strong belief in the firm’s rationality the consumer believes the firm expects her to $Buy$. Having received the move the consumer updates her expected payoff to at least 0 achievable by ending the game with $\neg Buy$. Should the consumer choose to $Buy$ the most she can signal to the worker that she expects is a payoff of at least 0. When it becomes the worker’s move he believes that the consumer believes he will choose $e_h$ with
probability of at least $\frac{1}{3}$. With this adaptation to the modeling of his simple guilt preferences, the worker pays attention to the consumer’s expectation held at the time of her move when assessing guilt, an anticipated payoff of 0. Seeing that the consumer receives a payoff that is below this expected value when low effort is chosen and one that exceeds this expected value when high effort is selected the worker selects high effort when $2 - \theta_{wc}(0) \leq 1$, or when $\theta_{wc} \geq 1$. Given common strong belief in rationality the consumer must also figure this out. For intermediate values of $\theta_{wc} \in \left[\frac{1}{3}, 1\right)$, the prediction is ambiguous. However, $[\text{Out}, \neg \text{Buy}, e_l]$ is not rationalizable in $\Gamma_3(x)$ for high values of $\theta_{wc}$ with modified simple guilt preferences.

This result demonstrates that when extensive form rationalizability is applied to the game in which the worker is endowed with BD (2007) simple guilt preferences adapted to be sensitive to beliefs held at the time of the consumer’s move, both the modeling of guilt aversion and the prediction accord with the intuition from the vignette. After all, Bill assesses his degree of guilt relative to the expectation held by the consumer at the time of her purchase and his predicted behavior is driven by this conditional belief. When Bill is sufficiently guilt averse he will exert high effort so as to avoid disappointing the consumer.

Taken together these predictions of SE and extensive form rationalizability align with the predictions in the BD (2009) trust game. The findings in Claim 1 show that there are multiple SE and the findings in Claim 2 show that there is a unique rationalizable outcome. Importantly, Claim 2 has displayed the intuition that one might hope to capture. In particular the guilt averse worker lives up to his expectation of the consumer’s belief at the time of her move. When the worker is sufficiently guilt averse he will exert high effort whether he initially expected the consumer would make a purchase or not.

5 Concluding remarks

In the previous section I consider an adaptation to BD (2007) preferences and demonstrate that the resulting model matches the psychological and economic intuition from the vignette. Crucially, the worker’s assessment of the degree to which he may let down the consumer is measured relative to the beliefs held by the consumer at the time of her move. In this case this corresponds to Bill assessing his potential disappointment of the young customer relative the the beliefs he holds about her expected payoff at the time she purchases the cotton candy. If Bill is sufficiently sensitive to guilt, the firm is able to count on his high effort selection even in the absence of monitoring. Presumably the firm may simply need to incur some fixed cost to induce guilt, for instance, making arrangements so that the worker and consumer interact when the good is transferred. As long as Bill is guilt averse and knows he will need to see the expression on the customer’s face, he will avoid shirking.
When faced with the problem of making sure that the worker exerts sufficiently high effort to produce a quality good that will be purchased the firm has a few options available. However, at the end of the day the firm may prefer to play $\Gamma_3$ to $\Gamma_2$. That is, the firm may prefer to induce guilt in the worker over the option of monitoring effort. This will certainly be the case whenever both are available and when $\mu > 0$, since the firm will have a strictly higher payoff when it allows the worker’s guilt aversion to bring about the efficient outcome.

In this paper I work toward accomplishing two objectives. First, I incorporate guilt aversion into a labor game in a way not yet considered by labor economists. I show that to the extent that a worker is sensitive to guilt higher efforts may be induced to avoid disappointing consumers. Second, I demonstrate that one must take care when directly applying the Battigalli & Dufwenberg (2007) model of simple guilt preferences. The reliance on initial beliefs to determine guilt may not appropriately capture the relevant intuition. I have not proposed a way to modify BD (2007) to suit general games and I suspect that it is best for individual applied economists to make changes as needed for their particular contexts. Further research may be warranted in this regard.
6 References


