CESifo WORKING PAPERS

8992
2021

Original Version: April 2021
This Version: April 2023

Classical and Belief-Based Gift Exchange Models: Theory and Evidence

Sanjit Dhami, Mengxing Wei, Ali al-Nowaihi

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 8992*

Classical and Belief-Based Gift Exchange Models: Theory and Evidence

Abstract

We derive and test the predictions of three competing models of gift exchange: Classical (CGE); Augmented (AGE) based on unexpected wage surprises; and Belief-based (BGE) that uses belief hierarchies to formally model reciprocity and guilt-aversion. Following Akerlof (1982), we also introduce signals of the typical wage, θ_w, and effort level, θ_e, in similar firms. We examine the worker's optimal effort in response to exogenous variation in the wage, w, the signals θ_w, θ_e, and a signal of the firm's expectations of effort from the worker, s. All three models predict gift exchange, however, the predictions of the AGE and the CGE models with respect to θ_w, θ_e, and s, are rejected. The BGE model successfully explains the data in all these respects. Gift exchange is underpinned by guilt-aversion. We also provide novel empirical evidence of first order stochastic dominance of first and second order beliefs.

JEL-Codes: D010, D910.

Keywords: gift exchange, reciprocity, guilt-aversion, psychological game theory, belief-based models, industry wage and effort norms.

Sanjit Dhami
University of Leicester
Department of Economics, School of Business
United Kingdom - Leicester LE2 1RQ
sd106@le.ac.uk

*Mengxing Wei**
School of Economics
Nankai University
China - 300071, Tianjin
mengxing.wei@hotmail.com

Ali al-Nowaihi
Department of Economics, School of Business, University of Leicester
United Kingdom - Leicester LE2 1RQ
aa10@le.ac.uk

*corresponding author

April 13, 2023

1 Introduction

In a *gift exchange game*, if a worker is offered a binding wage, w, by a firm in excess of the worker's outside option, then the worker responds by putting in greater non-binding effort, e, independent of reputational concerns, information asymmetries, and market imperfection (Mauss, 1924; Gouldner, 1960; Blau, 1964; Akerlof, 1982). By contrast, if the worker was purely self-regarding and lacked reciprocity considerations, then the worker would fully shirk. These results have played an important part in establishing the roles of *prosociality* and *reciprocity* in economics.

In seminal experiments, Fehr et al. (1993, 1998) and Fehr et al. (1997) demonstrated *conditional reciprocity* to a gift, after controlling for potential confounds, and derived far reaching implications, including those for the efficiency of competitive equilibria.[1] It has also led to important advances in contract theory, labour economics, and macroeconomic models (Dhami, 2019, Vol. 2). The gift exchange phenomenon survives large stakes experiments (Fehr et al., 2002). While most experiments have been done in the lab, gift exchange has also been documented with field data (Falk, 2007; Gneezy and List, 2006; Fehr et al., 2009; Bellemare and Shearer, 2007) and it is likely to have an evolutionary basis (Bowles and Gintis, 2011).

Yet the underlying motivations behind gift exchange deserve further exploration. The aim of our paper is to theoretically formulate some of the leading alternative explanations, derive testable predictions, and then subject them to stringent empirical tests in carefully controlled experiments.

1.1 Classical gift exchange (CGE)

The anthropological explanation for gift exchange draws on the innate desire of humans to respond positively to a gift by another, irrespective of any other considerations.[2] The classical gift exchange experiments were motivated by the seminal work of Akerlof (1982). Akerlof (1982) also allows for a signal of wages offered in similar firms (or *exogenous wage norms*), θ_w, to influence the worker's effort in a firm. But a signal of effort levels in similar firms in the industry (or *exogenous effort norms*), θ_e, does not play a central role in his formal analysis (see equation (6) on p.557 in Akerlof, 1982). Instead, Akerlof (1982) argued that *endogenous effort norms* within a firm that arise as part of the culture within a firm, may also influence effort. Yet, endogenous wage norms might be strongly influenced by exogenous wage norms. We shall consider only 'exogenous wage' and 'exogenous effort' norms, hence, we drop the prefix 'exogenous'.[3]

Akerlof (1982) uses 'norms' to refer to expectations of behaviors/actions of players that might be determined by norms of fairness or reciprocity, and uses a competitive general equilibrium model. Our interest is in the behavior of workers who are matched with a particular firm, as in typical gift exchange experiments, while introducing exogenous variation in these norms or expectations. In

[1] In a theoretical model, Netzer and Schmutzler (2014) show that when the firm's wage increase cannot be interpreted as a kind intention, then the worker may not respond with a positive gift exchange. Dufwenberg and Kirchsteiger (2019) show that these results depend on the definition of efficiency that is used.

[2] Malmendier and Schmidt (2017) nicely capture the anthropological insights behind gift exchange: "Our evidence suggests that a gift triggers an obligation to repay, independently of the intentions of the gift giver and the distributional consequences. It seems to create a bond between gift giver and recipient, in line with a large anthropological and sociological literature on gifts creating an obligation to reciprocate."

[3] It would appear that more foundational work is required on the nature of endogenous norms before they can be formalized in theoretical models.

this paper, we are not interested in the question of how these norms are determined or how they evolve; this will require a different paper. For this reason, we do not use the word 'norm' in our experimental instruction at all. In our experiments, θ_w is referred to as 'the typical wage paid in similar firms in the industry' and θ_e, as the 'typical effort exerted by workers in similar firms in the industry.' Thus, our use of the term 'norms' in the paper is in the spirit of Akerlof (1982), but it should not be confused with the usage of this term in the more recent, and separate, literature on social norms (Dhami, 2019, Section 5.7).

We use the acronym *CGE model* (Classical Gift Exchange model) for a model where the material utility of a worker, u, is augmented to include, in Akerlof's (1982) terminology, signals of wage norms, θ_w, and effort norms, θ_e. In the experimental literature, θ_w is sometimes interpreted as the outside option of the worker, and θ_e is interpreted as some notion of reasonable or perfunctory effort by the worker.

1.2 Augmented gift exchange (AGE)

Recent developments augment the CGE model to incorporate the first order beliefs/expectations of the players (Malmendier and Schmidt, 2017). If the worker receives an unexpected positive wage surprise, i.e., a wage in excess of wage expectations (*first order beliefs of the workers*), then the worker exerts higher effort.[4] A negative wage surprise may induce the worker to put in lower effort. The role of unexpected wage surprises has been studied in several experiments (Gneezy and List, 2006; Kube et al., 2013). Effort reciprocity on the part of the worker is found to be *gift-dependent*, but *intentions-independent* (Malmendier and Schmidt, 2017). We call a model that includes a wage surprise term, yet is intentions-independent, an *AGE model* (Augmented Gift Exchange model).

1.3 Belief-based gift exchange (BGE)

First order beliefs are central in the AGE model. However, we classify *belief-based models* as those which require the essential use of *belief hierarchies*, i.e., beliefs about beliefs. We only need to use *second order beliefs*, which are beliefs about the first order beliefs of the other player.[5] This allows for a rigorous formulation of the emotions behind individual actions and enables us to obtain new and testable implications that can discriminate between the alternative models.

Second order beliefs play an important role in determining *conditional reciprocity* and *guilt*.[6] Conditional reciprocity is known to underpin gift exchange behavior. In belief-based models, *reciprocity* depends on the beliefs workers have about the kindness intentions of the firm (workers' second order beliefs) as in models of *intentions-dependent reciprocity*. Workers may feel *guilty* if they let down the firm's expectations of effort from them. This requires workers to guess what effort the firm expects from them (the worker's second order beliefs about the first order beliefs of

[4] Englmaier and Leider (2012) also propose a similar formulation within a principal-agent problem. A similar transmission channel was used in macroeconomic models starting in the late 1960s to model the supply side of macroeconomic models, e.g., the Lucas and Rapping (1969) supply curve.

[5] Dhami et al. (2019) show, for instance, that fourth order beliefs play a role in determining public goods contributions in public goods games.

[6] With the exception of Dhami et al. (2019), the literature typically models conditional reciprocity and simple guilt separately (e.g., Dufwenberg et al., 2011, who use public goods games). Here we model these two aspects simultaneously.

the firm).

Guilt and conditional reciprocity are formally modeled in *psychological game theory*, where belief hierarchies (i) directly enter into the utility functions of players, (ii) and these beliefs can be endogenous (Geanakoplos et al., 1989; Rabin, 1993; Battigalli and Dufwenberg, 2009, 2022). A substantial literature now shows that players derive utility from reciprocity and disutility from guilt (Bowles and Gintis, 2011; Henrich, 2016; and for surveys, see Battigalli and Dufwenberg, 2022 and Dhami, 2020, Volume 4, sections 2.5, 3.6). Conditional reciprocity in psychological game theory was introduced by Rabin (1993), extended to extensive form games by Dufwenberg and Kirchsteiger (2004), and to an even more general framework by Battigalli and Dufwenberg (2009). In psychological game theory, guilt was introduced by Battigalli and Dufwenberg (2007); we are mainly interested in what they term as *simple guilt*.[7] Guilt also lies behind many other explanations of human behavior including moral norms and the norm of reciprocity (Bicchieri, 2006; Elster, 2011). The applications have increased considerably in recent years (Battigalli and Dufwenberg, 2022 and Dhami, 2020, Volume 4).

1.4 Experiments and main findings

We begin by deriving the theoretical predictions of the three competing models CGE, AGE, and BGE. These take the form of comparative static results with respect to the parameters w, θ_w, θ_e, s. The first set of experiments were conducted in the lab during March-May, 2019, in China with 240 university students. A second set of experiments were conducted in June-July 2020 in China with 58 subjects on a similar subject pool. In total, and using the strategy method, across all sub-treatments we collected $12,144$ data points for decisions made by firms and workers. We conducted a third set of experiments in March 2022, with 182 students, at Nankai University in order to test for the first order stochastic dominance of first order and second order beliefs.

All three models, CGE, AGE, and BGE, predict an increase in the worker's optimal effort in response to an increase in w (gift exchange), in line with Akerlof (1982) and our empirical evidence. Hence, gift exchange alone cannot discriminate among the models. A key role in distinguishing between the three models is played by the three signals, θ_w (industry wage norm), θ_e (industry effort norm), and s (firm's expectation of effort from its matched worker). Our empirical evidence shows that effort increases in response to an increase in all three signals θ_w, θ_e, s.

Only the BGE model correctly predicts an increase in effort when θ_w, θ_e, s increase. The BGE model correctly predicts an increase in effort when the worker receives a high signal of the firm's effort expectation (high value of s). The reason is that letting down the firm's expectations induces 'guilt' in the worker, a channel that is absent in the CGE and the AGE models. The signal s also plays no role in Akerlof (1982), yet arguably an important part of any employment relation is that firms convey their effort expectations to workers through a variety of means, even if these expectations are non-binding.

[7] Another form of guilt aversion, *guilt from blame*, requires the formation of third and fourth order beliefs (Battigalli and Dufwenberg, 2007). For applications and the empirical evidence, see Khalmetski et al. (2015) and Dhami et al. (2019). Our characterization of guilt-aversion accords well with survey measures of guilt used in psychology such as those embodied in TOSCA3 (Bellemare et al., 2019). Falk and Fischbacher (2006) develop a theoretical framework to explain positive gift giving in a psychological game theory model by appealing to the reciprocity explanation, but they do not consider guilt.

The comparative static effects of θ_w, θ_e, s on the worker's effort in the CGE and the AGE models are rejected by the evidence. Having already considered the comparative static effect of s, we now briefly explain the intuition behind the comparative static results with respect to θ_w, θ_e in the three models.

In the CGE model, the higher is the wage norm θ_w, the lower is the size of the perceived gift (difference between the wage paid and the wage norm) by the matched firm. Hence, optimal effort is reduced (contrary to our evidence). The effort norm signal, θ_e, has no effect on optimal effort in the CGE model (as in Akerlof, 1982) due to the linearity of the objective function. In the AGE model, the main transmission channel is unexpected wage changes. An increase in the wage norm, θ_w, increases the wage expectation of the worker, reducing the size of unexpected wage surprises and, hence, reducing optimal effort; this is contrary to our evidence. We provide a direct test for this transmission mechanism in Section 12 that confirms the mechanism.

In the BGE model, an increase in either θ_w or θ_e, increases the likelihood that the firm expects a higher effort from the worker, as perceived by the worker. If the effort norm, θ_e, in the industry is high, then the worker may reasonably believe that the matched firm also expects a high effort from its workers. If the wage norm, θ_w, is high, then given that paying higher wages is costly to the firm, workers should reasonably expect that this creates an expectation of a higher effort. Our Assumption A4 of Section 5 formulates these insights. A consequence of this assumption is that the distribution of second order beliefs of the worker has the first order stochastic dominance property with respect to the parameters θ_w, θ_e, s (Corollary 2). Our empirical results in Section 12 directly verify this property of beliefs, which is another novel contribution of our paper. Guilt averse workers must then increase their optimal effort so as not to fall below the firm's expectation.

We believe that our treatment of the gift exchange experiments is arguably closer in spirit, albeit not identical, to Akerlof (1982). A consideration of wage changes alone, without exogenous variation in the three signals θ_w, θ_e, s is unable to differentiate between the three models CGE, AGE, and BGE. An advantage of developing a proper and rigorous theoretical framework is that it enables us to separate the predictions of the three models and guides us in the design of the different treatments in our experiments. In this sense, we contribute towards a greater overall understanding of gift exchange.

1.5 Relation to the literature

The gift exchange relation is also predicted by other models such as other-regarding preferences (e.g., *inequity aversion*, Fehr and Schmidt, 1999) and *type-based reciprocity* (Levine, 1998). Malmendier and Schmidt (2017) show that in their setup with third party advice, the AGE model explains their data better than either the inequity aversion model or type-based reciprocity. Models of inequity aversion and type-based reciprocity do not formally include beliefs into the utility functions of the player. Hence, they are unlikely to be able to explain the variation in our data with respect to the signals, θ_w, θ_e, s.[8]

[8]This is not a criticism of these models, which were developed to explain a different set of stylized facts. Furthermore, the inequity aversion model can explain the existence and efficacy of punishments in the gift exchange game which the CGE and the AGE models cannot. However, invoking emotions such as anger, frustration, and disappointment (see Battigalli and Dufwenberg, 2022 for the formal definitions) may also be able to provide an

The early empirical literature on guilt-aversion used *direct belief elicitation* and concluded that guilt aversion was important.[9] This required asking players to directly state their beliefs about another player's first order beliefs. Such *direct belief elicitation* of second order beliefs is, however, subject to the *false consensus effect* (Ellingsen et al., 2010); namely, players assign their own beliefs to others when asked to guess the beliefs of other. This confound invalidates tests based on direct elicitation methods. We follow an amended version of the *induced beliefs design* of Ellingsen et al. (2010) that eliminates the false consensus effect; see Section 7 below for a discussion. We are the first to apply the induced beliefs design to the gift exchange game. The induced beliefs design has also found substantial evidence for the guilt aversion channel in a range of phenomena.[10]

There is a sizeable literature on gift exchange in the lab and the field that considers different features of gift exchange. This includes piece rates (DellaVigna et al., 2020); effort response to positive and negative wage surprises relative to an originally promised wage (Gneezy and List, 2006; Kube et al., 2013); in-kind gifts (Kube et al., 2012); tenure effects on gift exchange (Bellemare and Shearer, 2009); and social comparisons and gift exchange (Abeler et al., 2010; Gächter et al., 2012). A more complete survey can be found in Dhami (2019, Vol. 2, Section 1.3) and for a survey of experiments using the piece rate design, see DellaVigna et al. (2020). However, that literature does not apply belief hierarchies to the gift exchange game. Hence, in that literature, guilt, reciprocity, and the effort response to the signals θ_w, θ_e, s, do not rely on a formal and rigorous belief-based framework.

The literature on peer effects also identifies the role of guilt and shame in influencing the effort exerted by economic agents in team environments (Kandel and Lazear, 1992; Mas and Moretti, 2009). However, this literature does not formally model the beliefs of the players, but see Dhami et al. (2020) who provide the belief-based foundations and the empirical evidence. For a related literature on the effects of changes in the outside option of workers in the form of wages, on the worker's optimal effort, see Charness and Kuhn (2007), Abeler et al. (2010), and Bejarano et al. (2021). In particular, Bejarano et al. (2021) show that in times of economic instability, when firms and workers may differ in their perceived fairness of the outside option, the perceived gift embedded in a wage increase may fall, weakening the gift exchange relation.[11] Our results provide some of the microfoundations for this literature.

alternative explanation for the existence of punishments within the domain of our BGE model.

[9] Since our study does not use the direct belief elicitation method, we omit the large number of references. The necessary references can be found in Battigalli and Dufwenberg (2022). See also the literature survey in Cartwright (2019) and the references in Dhami et al. (2019).

[10] For instance, guilt aversion explains greater dictator giving when the receiver expects more (Khalmetski et al., 2015; Hauge, 2016); has a critical influence on public good contributions in public goods games (Dhami et al., 2019); influences embezzlements decisions in an embezzlement game (Attanasi et al. 2019); influences the prosociality of choices in advisor-consumer experiments (Inderst et al., 2019); may potentially explain the decision to keep promises (Di Bartholomew et al., 2018); explains the participation decision in public goods games (Patel and Smith, 2019); influences dictator decisions in dictator games so long as the receiver does not expect more than half the surplus (Balafoutas and Fornwagner, 2017); and explains the structure of microfinance contracts (Dhami et al., 2020). For surveys, see Battigalli and Dufwenberg (2022), Cartwright (2019), and Dhami (2020, Vol. 4). For general issues in belief elicitation that go beyond psychological game theory, see Schotter and Trevino (2014).

[11] We are very grateful to Referee 2 for alerting us to this literature.

1.6 Plan of the paper

Section 2 describes the basic set-up of our model: Preferences and technology, and the formation and updating of beliefs. Sections 3 and 4 describe, respectively, the CGE and the AGE models. Section 5 describes the BGE model. In each case, we also derive the theoretical predictions of the model. Section 6 summarizes the comparative static results of the models for ease of future reference. Section 7 describes the experimental design. Section 8 compares the predictions of the three models with respect to the empirical evidence. Section 9 considers the determinants of effort in a Tobit censored regression analysis. Section 10 discusses the results. Section 11 discusses the results of the post-experimental survey. Section 12 describes the results of a fresh set of experiments designed to test for our assumptions on first order and second order beliefs (assumptions A3 and A4, respectively). Section 13 summarizes and concludes. All proofs are contained in the Appendix. A supplementary section contains the detailed experimental instructions.

2 The Model

We describe our model below. Our results can be generalized, but we suppress generality to keep the model as close as possible to our experimental design.

2.1 Preferences and technology

Consider a standard gift exchange game between one firm (F) and one worker (W). The production function of the firm is $y(e) = e$, where $e \in [0,1]$ is the effort level of the worker. Each unit of output can be sold by the firm at an exogenously given product price 100. Hence, the *profit* of the firm is $\pi(e,w) = 100e - w$, where $w \in [0,100]$ is the contractually binding wage paid to the worker, i.e., it is not conditional on the effort level of the worker. In keeping with the ethical requirement in experiments that subjects in their roles as firms are never out of pocket, we add a fixed number, $\kappa \geq 0$, to the profit function[12]. Thus, we adopt the profit function

$$\pi(e,w) = 100e - w + \kappa;\ w \in [0,100],\ e \in [0,1]. \tag{2.1}$$

The presence of the fixed number κ has no effect on any of our results.

The *material utility* of the worker is

$$u(e,w) = w - 20e^2;\ w \in [0,100],\ e \in [0,1], \tag{2.2}$$

where $20e^2$ is the disutility of work to the worker.

2.2 Beliefs and sequence of moves

The generic gift exchange game has two stages that we describe below, following the description of initial beliefs.

[12]For instance, evaluated at $w = 40$ and $e = 0.4$, we get $\pi(40, 0.4) > 0$ when $\kappa = 40$. In case the profits are negative, the corresponding amount is subtracted from the participation fee of 20 Yuan. However, on net, subjects are never out of pocket in any scenario in our experiment.

2.2.1 Initial beliefs

At the start of the experiment, once the structure of the gift exchange game is revealed to the subjects (as in subsection 2.1 above) *but before any of the subjects takes any action*, we assume that the subjects have initial beliefs. We will mainly be interested in cumulative distribution functions. *First order beliefs*: These are beliefs about the choices that one expects other players to make and are superscripted with '1'. In particular, $P_F^1(e)$ is the probability with which the firm believes the worker will exert an effort less than or equal to $e \in [0,1]$. Analogously, $P_W^1(w)$ is the probability with which the worker believes the firm will offer a wage less than or equal to $w \in [0,100]$.[13]

Second order beliefs: These are beliefs of the players about the first order beliefs of other players; we superscript these with a '2'. $P_W^2(e)$ is the probability with which the worker believes that the firm expects an effort level less than or equal to e from the worker. We do not need to specify the second order beliefs of the firm, P_F^2, as they play no role in our analysis.

2.2.2 Stage 1 (Firms: Belief and wage setting)

The sequence of moves in Stage 1 is as follows.

1. The experimenter conveys to the firm an effort signal, $\theta_e \in [0,1]$, and a wage signal $\theta_w \in [0,100]$ of, respectively, the effort and wage levels in similar firms.
2. After observing these signals the firm updates its initial belief, $P_F^1(e)$, to $P_F^1(e \mid \theta_w, \theta_e)$.
3. The experimenter elicits from the firm a signal, $s \in [0,1]$, of the firm's first order belief, $P_F^1(e \mid \theta_w, \theta_e)$, about the effort level, e, expected from the worker. The signal s is a point estimate made by the firm. For example, s could be the mean, median, or mode of $P_F^1(e|\theta_w, \theta_e)$.
4. The firm chooses the contractible wage level $w \in [0,100]$ for the worker.

2.2.3 Stage 2 (Workers: Beliefs and choice of effort)

Workers are divided into two groups, uninformed workers (Treatment 1) and informed workers (Treatment 2). The distinction between the two groups is that the experimenter communicates the signal, s (taken from Stage 1), as a *private signal*, to the informed group of workers (Treatment 2), but not to the uninformed group of workers. Half the workers are in each treatment. Thus, we can differentiate the information sets of the two types of workers in the following definition.

Definition 1 *: Define Γ_j to be the information set of the worker of type $j = I, N$, where I denotes an 'informed worker' who knows the signal $s \in [0,1]$ of the firm's expected effort level, e, and N denotes an uninformed worker who does not know the signal s. In particular, and restricting attention to the three signals, θ_w, θ_e, s (i.e., excluding the wage, w), we have*

$$\Gamma_I = \{\theta_w, \theta_e, s\} \text{ and } \Gamma_N = \{\theta_w, \theta_e\}. \tag{2.3}$$

[13] A simpler model might have used point beliefs. However, a distribution of beliefs allows for the more realistic situation of underlying, but unmodelled, uncertainty about the actions of others, and allows for a more elegant derivation of the results.

The sequence of moves in Stage 2 is as follows.

1. All workers observe the signals of industry effort and wage norms, θ_e and θ_w; these are the same signals as observed by the firm. Informed workers are informed of the signal, s, while uninformed workers are not informed of the signal, s.

2. Using the signals, θ_w, θ_e, s, workers update their initial first order beliefs from $P_W^1(w)$ to $P_W^1(w \mid \Gamma_j)$, and their second order beliefs from $P_W^2(e)$ to $P_W^2(e \mid \Gamma_j)$, $j = I, N$.

3. The experimenter reveals to the worker the contractible wage level $w \in [0, 100]$ chosen by the firm for the worker in Stage 1.

4. The worker then chooses his/her effort level $e \in [0, 1]$.

This concludes Stage 2.

The following tables summarize the beliefs of the firm and the worker. As noted above, the second order beliefs of the firm, P_F^2, play no role in our analysis, hence, we omit them from the tables.

Initial beliefs	Firm	Worker
First order beliefs	$P_F^1(e)$	$P_W^1(w)$
Second order beliefs	-	$P_W^2(e)$

Table 1: Initial beliefs of the firm and workers.

Updated beliefs	Firm	Uninformed worker	Informed worker
First order beliefs	$P_F^1(e \mid \theta_w, \theta_e)$	$P_W^1(w \mid \theta_w, \theta_e)$	$P_W^1(w \mid \theta_w, \theta_e, s)$
Second order beliefs	-	$P_W^2(e \mid \theta_w, \theta_e)$	$P_W^2(e \mid \theta_w, \theta_e, s)$

Table 2: Updated beliefs of the firm and workers.

2.3 Technical assumptions on beliefs

We are primarily interested in the cumulative belief distribution functions of workers. Assumptions A1 and A2 are purely technical assumptions that facilitate our exposition.

Assumption A1 *(Continuity): $P_W^1(w)$ and $P_W^2(e)$ are continuous functions of w and e, respectively, hence integrable.*

Assumption A2 *(Differentiability): $P_W^1(w \mid \Gamma_j)$ and $P_W^2(e \mid \Gamma_j)$, $j = I, N$, are continuously differentiable with respect to $\gamma \in \Gamma_j$.*

We shall make two further assumptions on the belief distributions that have substantive economic content: Assumption A3 in Section 4 (the AGE model) and Assumption A4 in Section 5 (the BGE model). We also provide direct empirical tests of these assumptions.

2.4 A note on best response to beliefs

The main solution concept in psychological game theory is a variant of *sequential equilibrium*.[14] Actions are optimal given beliefs, and beliefs of all orders are correct given actions (the assumption of *rational expectations*). This typically gives rise to multiple equilibria that make testing of the theory difficult (Rabin, 1993; Fehr and Schmidt, 2006; Malmendier and Schmidt, 2017). Furthermore, and contrary to the rational expectations assumption, there is much evidence for disequilibrium beliefs in experimental games.[15] Bellemare et al. (2011) show persuasively that actions taken by the players, their first order beliefs, and their second order beliefs, are not mutually consistent with each other, violating the rational expectations assumption. Polonio and Coricelli (2018) use eye tracking evidence to locate the reason for the lack of consistency between actions and beliefs.

In light of the evidence, it is unjustified to use the assumption that players have correct beliefs and that these are mutually consistent with equilibrium actions (Battigalli and Dufwenberg, 2022). Hence, we do not impose the condition that beliefs of all orders are consistent with the equilibrium actions but, instead, assume that players play a *best response to their beliefs* (Battigalli and Dufwenberg, 2022). In our simple model, the solution is likely to be similar to a rationalizable equilibrium (Pearce, 1984). The predictions of these models, particularly that use psychological game theory, match well with the evidence (Khalmetski et al., 2015; Dhami et al., 2019; Dhami et al., 2020). The problem of multiple equilibria is also avoided in our setup.

3 Classical gift exchange (CGE)

In the classical gift exchange game, for a given wage, w, the worker chooses the optimal effort, $e = e^*$, to maximize the following utility function (recall that the index $j = I, N$ denotes 'informed workers' who receive the signal s and 'uninformed workers' who do not receive this signal).

$$W\left(e; w, \Gamma_j\right) = w - 20e^2 + \beta\left(w - \theta_w\right)\left(e - \theta_e\right);\ j = I, N,\ \beta > 0. \tag{3.1}$$

The first two terms on the RHS of (3.1) give the worker's material payoff. The third term, $\beta\left(w - \theta_w\right)\left(e - \theta_e\right)$, gives the 'reciprocity payoff.' Suppose that the firm pays the worker a wage, w, above (below) the signal of the wage in similar firms, θ_w. Then the worker increases utility by exerting effort, e, above (below) the signal of effort in similar firms, θ_e. To quote from Akerlof (1982, p. 544): *On the worker's side, the "gift" given is work in excess of the minimum work standard; and on the firm's side the "gift" given is wages in excess of what these women could receive if they left their current jobs.* We believe that (3.1) is closer to Akerlof's original formulation of gift exchange.[16]

The third term in (3.1) is the essence of gift exchange in this *intentions-independent* frame-

[14] For a fuller and sophisticated discussion of these issues, which is outside the scope of our paper, see Battigalli et al. (2019) and Battigalli and Dufwenberg (2022).

[15] Disequilibrium beliefs are incorporated into several prominent behavioral models such as the level-k model, the cognitive hierarchy model, evidential equilibrium, and analogy based equilibria (Dhami, 2020, Vol. 4).

[16] In many lab experimental gift exchange studies, θ_w is set as the outside option of the worker, and typically normalized to zero, and θ_e is not taken into account. Hence, the third term in (3.1) becomes βwe.

work.[17] Note that the signal, s, of the firm's expectation of effort from the worker, plays no role in the CGE model because it lacks a formulation of emotions such as guilt.

Proposition 1 *: The worker's utility (3.1) has a unique maximum, $e^*(w, \Gamma_j)$. It has the following properties:*

(a) If $w \leq \theta_w$, then $e^*(w, \Gamma_j) = 0$.

(b) If $\theta_w < w < \frac{40}{\beta} + \theta_w$, then $e^*(w, \Gamma_j) = \frac{\beta}{40}(w - \theta_w)$ and, consequently:

(bi) $\frac{\partial e^*(w,\Gamma_j)}{\partial w} > 0$,

(bii) $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta_w} < 0$,

(biii) $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta_e} = \frac{\partial e^*(w,\Gamma_j)}{\partial s} = 0$.

(c) If $w \geq \frac{40}{\beta} + \theta_w$, then $e^*(w, \Gamma_j) = 1$.

Discussion of Proposition 1: From (bi) optimal effort is increasing in the wage rate, w (gift exchange). However, from (bii), it is decreasing in the signal of the typical wage in similar firms, θ_w; the reason is that a higher θ_w reduces the gift, $w - \theta_w$, as perceived by the worker. This conclusion is unaltered if, as in typical gift exchange experiments, one interprets θ_w as the outside option of the worker. The CGE model does not take account of guilt, hence, s has no effect on the optimal effort level, e^*. The marginal benefit from an extra unit of effort is independent of θ_e, hence e^* is independent of θ_e. In sum, the CGE model predicts $\frac{\partial e^*(w,\Gamma_j)}{\partial w} > 0$, which is the essence of gift exchange, and this is consistent with our evidence. However, from (bii) it also predicts $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta_w} < 0$ and from (biii) $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta_e} = \frac{\partial e^*(w,\Gamma_j)}{\partial s} = 0$; both are rejected by our data.

4 Augmented gift exchange (AGE)

The modern literature on gift exchange retains the central gift exchange feature of the CGE model but incorporates the role of expectations even more explicitly, particularly through the channel of wage surprises (Gneezy and List, 2006; Kube et al., 2013; Malmendier and Schmidt, 2017). Malmendier and Schmidt (2017) write (p. 495): "... *the weight that player i attaches to the welfare of player j depends on the actions of j that affect i, relative to the expected behavior of j. A favorable act such as giving a gift strengthens the bond between the giver and the recipient, i.e., the weight of the giver's payoff in the recipient's utility function, and the recipient will reciprocate. The key difference to existing models of action-based reciprocity ... is the prediction that the more the favorable act exceeds expectation, the stronger the positive response.*"

Following these insights, in our *augmented gift exchange* (AGE) model, the utility of the worker arising from (positive or negative) wage surprises, is given by

$$V(e; w, \Gamma_j) = w - 20e^2 + \frac{\sigma}{100}\pi(e, w)\left[w - E(w \mid \Gamma_j)\right], \tag{4.1}$$

where $\sigma > 0$, $j = I, N$, and the wage expected by the worker, given the signals in Γ_j, is

$$E(w \mid \Gamma_j) = \int_{w=0}^{100} w dP_W^1(w \mid \Gamma_j). \tag{4.2}$$

[17] Modeling intentions requires the explicit use of beliefs about the perceived kindness of the other player as in the BGE model in Section 5.

The first two terms on the RHS of (4.1) give the material payoff of the worker. The third term is the reciprocity payoff to the worker based on the wage surprise term, $w - E(w \mid \Gamma_j)$, which can be positive or negative.[18] Thus, the firm's profits, $\pi(e, w)$, are internalized positively or negatively in the worker's utility function depending on the sign of the wage surprise term. Reciprocity in the AGE model arises from an innate desire to reciprocate a gift, irrespective of the other player's intentions, i.e., reciprocity is *gift-conditional* and *intentions-unconditional*. Finally, we can consider an even richer model in which we include on the RHS of (4.1), the third term on the RHS of (3.1), $\beta(w - \theta_w)(e - \theta_e)$. However given the linearity of this term, this does not change any of our results. The factor $\frac{1}{100}$ in (4.1) is introduced for computational convenience, without affecting any results.

Substituting from (2.1) into (4.1) we get for $j = I, N$,

$$V(e; w, \Gamma_j) = w - 20e^2 + \frac{\sigma}{100}(100e - w + \kappa)[w - E(w \mid \Gamma_j)]. \tag{4.3}$$

Let $\overline{\gamma} = 100$, if $\gamma = \theta_w$; and let $\overline{\gamma} = 1$, if $\gamma \in \{\theta_e, s\}$.[19]

Assumption A3 *We make the following assumption on the conditional expected wage of the worker, $E(w \mid \Gamma_j)$:*

$$\frac{\partial E(w \mid \Gamma_j)}{\partial \gamma} > 0 \text{ for all } e \in (0,1),\ \gamma \in \Gamma_j,\ \gamma \in (0, \overline{\gamma}),\ j = I, N.$$

Discussion of Assumption A3: From Assumption A3, it follows that, for any $\gamma \in \Gamma_j$, $j = I, N$, the ex-ante belief of the worker is that a higher value of γ increases the expected wage. If firms operate in an industry where the typical wage, θ_w, is high, the worker believes that the matched firm is also 'more likely' to offer, on average, a higher wage. Or, if the informed worker receives a higher signal, s, of the effort expected by the firm from the worker, then the worker believes that it is 'more likely' that the firm will, on average, pay a higher wage. Finally, if the firm operates in an industry in which the typical effort exerted by workers, θ_e, is high, then since putting in higher effort is more costly, the worker believes that it is 'more likely' that the matched firm will offer, on average, a higher wage to compensate for the cost of effort. Our direct empirical test of Assumption A3 strongly confirms it; see Section 12.

Proposition 2 *The worker's utility (4.3) has a unique maximum, $e^*(w, \Gamma_j)$. If Assumption A3 holds, then it has the following properties:*
(a) If $w \leq E(w \mid \Gamma_j)$, then $e^*(w, \Gamma_j) = 0$.
(b) If $E(w \mid \Gamma_j) < w < \frac{40}{\sigma} + E(w \mid \Gamma_j)$, then
(bi) $\frac{\partial e^*(w,\Gamma_j)}{\partial w} = \frac{\sigma}{40} > 0$,
(bii) $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta_w} < 0$.

[18] This term is identical to the formulation in Malmendier and Schmidt (2017), and may also be taken to correspond to what Akerlof (1982, p.557) terms as *endogenous* work norms within the firm.

[19] In order to define first order stochastic dominance (see, e.g., Assumption A3 below) we need to define the domains of the relevant variables in the information sets Γ_j, $j = I, N$ of the two types of workers. Since wage $w \in [0, 100]$, thus, the signal of wage in similar firms $\theta_w \in [0, 100]$, hence the upper limit $\overline{\gamma} = 100$. Since $e \in [0, 1]$, the signals of effort in similar firms and the signal of the firm's first order effort expectations, $\theta_e, s \in [0, 1]$, hence, in this case $\overline{\gamma} = 1$.

(biii) $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta e} < 0$.

(biv) $\frac{\partial e^*(w,\Gamma_N)}{\partial s} = 0, \frac{\partial e^*(w,\Gamma_I)}{\partial s} < 0$

(c) If $w \geq \frac{40}{\sigma} + E(w \mid \Gamma_j)$, *then* $e^*(w,\Gamma_j) = 1$.

Discussion of Proposition 2: From (bi) we see that, in the AGE model, $\frac{\partial e^*(w,\Gamma_j)}{\partial w} > 0$, which captures the essence of gift exchange, and is in line with our evidence. However, from (bii), (biii), and (biv) for $j = I$, the predicted optimal effort in the AGE model is decreasing in the typical wage in similar firms, θ_w, the typical effort in similar firms, θ_e, and the signal, s, of the firm's expectation of effort from the informed worker. From Assumption A3, each of the three signals, θ_w, θ_e, and s for the informed worker, increases the expected wage of the worker. In turn, this reduces the wage surprise, and the marginal benefit of an extra unit of effort (the relevant first order conditions can be seen in the proof of Proposition 2). The predictions in (bii), (biii), and (biv) are inconsistent with our empirical evidence. The BGE model that we consider in Section 5 makes the opposite predictions and these are supported by the evidence.

5 A belief-based model of gift exchange (BGE)

As noted in the introduction, extensive evidence is consistent with the central predictions of models of psychological game theory. These models allow *beliefs* and *beliefs about beliefs* to directly enter into the utility function of individuals enabling the formal modelling of reciprocity and guilt. We call such models, when applied to the gift exchange game, *belief-based gift exchange* (BGE) models. In this case, the utility of the worker is

$$U(e,w,\Gamma_j) = w - 20e^2 + \frac{\lambda_R}{100} R(e,w,\Gamma_j) - \lambda_G G(e,\Gamma_j) + \lambda_S S(e,\Gamma_j), \tag{5.1}$$

where

$$\lambda_R > 0,\ \lambda_G > \lambda_S \geq 0, \tag{5.2}$$

are weights that measure the relative strengths of the associated terms. Γ_j is given by Definition 1. The factor $\frac{1}{100}$ in (5.1) is used for computational convenience, without affecting any results. The parameter restrictions in (5.2) are sufficient for U to be a strictly concave function of e. The BGE model augments material utility, $w - 20e^2$, with three additional terms based on R (reciprocity), G (guilt), and S (surprise). We describe these below.

Guilt-aversion (G): In (5.1), G is the worker's *guilt function*. Following Battigalli and Dufwenberg (2007), the guilt function of the worker (also termed as *simple guilt*) is:

$$G(e,\Gamma_j) = \int_{x=e}^{x=1} (x-e)\, dP_W^2(x \mid \Gamma_j),\ j = I, N. \tag{5.3}$$

From (5.3), the worker faces disutility on account of guilt if actual effort, e, falls short of what the worker believes is the firm's expectation of the effort level. Since workers do not directly observe the firm's expectations (or the first order belief of the firm, $P_F^1(e \mid \Gamma_j)$), they use their own second order beliefs, $P_W^2(e \mid \Gamma_j)$, which are beliefs about $P_F^1(e \mid \Gamma_j)$.

Surprise-seeking (S): In (5.1), the function, S, gives the *elation* received by the worker from exceeding the perceived expectations of the firm (Khalmetski et al., 2015; Dhami et al. 2019):

$$S\left(e,\Gamma_j\right)=\int_{x=0}^{x=e}\left(e-x\right)dP_W^2\left(x\mid\Gamma_j\right),\ j=I,N. \tag{5.4}$$

Khalmetski et al. (2015) and Dhami et al. (2019) explicitly measure the relative sizes of λ_G,λ_S in (5.1) and find that guilt-aversion is the more important of the two emotions (for 70% of the subjects in the first study and for 95% of the subjects in the second study). Hence, the restriction $\lambda_G>\lambda_S\geq 0$ in (5.2). We shall refer to the combination $-\lambda_G G\left(e,\Gamma_j\right)+\lambda_S S\left(e,\Gamma_j\right)$ is (5.1) as the *guilt-aversion channel*.

Integrating (5.3) and (5.4) by parts, we get:

$$G\left(e,\Gamma_j\right) = 1-e-\int_{x=e}^{x=1}P_W^2\left(x\mid\Gamma_j\right)dx, \tag{5.5}$$

$$S\left(e,\Gamma_j\right) = \int_{x=0}^{x=e}P_W^2\left(x\mid\Gamma_j\right)dx,\ j=I,N. \tag{5.6}$$

Reciprocity (R): We apply the framework of Dufwenberg and Kirchsteiger (2004) for sequential games to specify the conditional reciprocity term $R\left(e,w,\Gamma_j\right)$ in (5.1).[20] In this framework,

$$R\left(e,w,\Gamma_j\right)=k_{WF}\left(e,w\right)\widehat{k}_{FW}\left(w,\Gamma_j\right), \tag{5.7}$$

where $k_{WF}\left(e,w\right)$ is the *kindness of the worker to the firm, as perceived by the worker* and $\widehat{k}_{FW}\left(w,\Gamma_j\right)$ is the *kindness of the firm to the worker*, also as *perceived by the worker*. If the firm is perceived to be kind ($\widehat{k}_{FW}>0$), then by reciprocating the kindness ($k_{WF}>0$), the worker increases utility as given in (5.1). Similarly, utility can be increased by reciprocating unkindness ($\widehat{k}_{FW}<0$) with unkindness ($k_{WF}<0$). We first state the basic concepts in Definition 2, below, and then explain them.

Definition 2 *(Dufwenberg and Kirchsteiger, 2004):*
(a) The equitable payoff of the firm, $\pi^E\left(w\right)$, as perceived by the worker, is
$\pi^E\left(w\right)=\mu\max\left\{\pi\left(e,w\right),\ e\in[0,1]\right\}+\left(1-\mu\right)\min\left\{\pi\left(e,w\right),\ e\in[0,1]\right\}$,
where $\pi\left(e,w\right)$ is given by (2.1) and $\mu\in[0,1]$.
(b) The kindness of the worker to the firm, $k_{WF}\left(e,w\right)$, as perceived by the worker, is
$k_{WF}\left(e,w\right)=\pi\left(e,w\right)-\pi^E\left(w\right)$.
(c) Let the expected utility of the worker $Eu\left(w,\Gamma_j\right)=\int_{e=0}^{1}u\left(e,w\right)dP_W^2\left(e\mid\Gamma_j\right)$, where $u\left(e,w\right)$ is given by (2.2), and $P_W^2\left(e\mid\Gamma_j\right)$ is the cumulative probability distribution of the second order belief of the worker.
(d) The equitable payoff to the worker, $u^E\left(w,\Gamma_j\right)$, as perceived by the worker, is
$u^E\left(w,\Gamma_j\right)=\nu\max\left\{Eu\left(w,\Gamma_j\right),\ w\in[0,100]\right\}+\left(1-\nu\right)\min\left\{Eu\left(w,\Gamma_j\right),\ w\in[0,100]\right\}$,
where $u\left(e,w\right)$ is given by (2.2) and $\nu\in[0,1]$.
(e) The kindness of the firm to the worker, $\widehat{k}_{FW}\left(w,\Gamma_j\right)$, as perceived by the worker, is

[20] For the relative merits of alternative methods of modeling kindness functions, such as those based on the seminal model by Rabin (1993), see Dufwenberg and Kirchsteiger (2019) and Dhami (2020; Vol. 4, Section 2.5). However, in our model, the two different methods do not yield substantively different results.

$\widehat{k}_{FW}(w, \Gamma_j) = Eu(w, \Gamma_j) - u^E(w, \Gamma_j)$.
(f) The worker's conditional reciprocity towards the firm, $R(e, w, \Gamma_j)$, is given in (5.7).

Discussion of Definition 2: The essential idea behind kindness functions is to compare the *expected payoff of a player* with the player's *equitable payoff* that captures some notion of a fair payoff. The equitable payoff is a weighted average of the maximum and the minimum payoffs that accrue to a player, from the actions taken by the other player. In Definition 2a, in Stage 2, after having observed the wage announced by the firm, the worker computes the equitable payoff of the firm, as perceived by the worker. It is a weighted average of the maximum and the minimum profits of the firm that arise from the actions of the worker. By choosing $e = 0$, the worker ensures profits of the firm are minimized, and by choosing $e = 1$, the profits are maximized. The kindness of the worker to the firm, in Definition 2b, is the difference between the expected and the equitable profits of the firm.

In computing the equitable payoff of the worker, as perceived by the worker, the worker needs to know the firm's expectations of the worker's effort. Since these expectations are unobserved, the worker forms second order beliefs about these expectations (Definition 2c). The action of the firm that leads to the maximum payoff for the worker is $w = 100$, and the action that leads to the minimum payoff is $w = 0$. Definition 2d computes the equitable payoff of the worker as a weighted average of the maximum and minimum payoffs. The kindness of the firm to the worker, as perceived by the worker, is computed in Definition 2e, as the difference between the expected payoff and the equitable payoff.

We can now use Definition 2 to state the expression for the reciprocity term for the worker, as perceived by the worker.

Proposition 3 *The worker's conditional reciprocity towards the firm, $R(e, w, \Gamma_j)$, is given by*

$$R(e, w, \Gamma_j) = 100(w - 100\nu)(e - \mu), \tag{5.8}$$

where $\mu \in [0, 1]$ and $\nu \in [0, 1]$ are as in Definition 2.

The parameter μ is positively associated with the firm's entitlement to an equitable payoff, *as perceived by the worker*. Hence, an increase in μ reduces the extent of kindness of the worker to the firm, k_{WF}, from the worker's actions. The parameter ν is positively associated with the worker's entitlement to an equitable payoff, *as perceived by the worker*, hence, an increase in ν reduces the extent of perceived kindness of the firm to the worker, $\widehat{k}_{FW}$, from the firm's actions. Dufwenberg and Kirchsteiger (2004) set $\mu = \nu = \frac{1}{2}$, however, we allow for the more general case. The interpretation of (5.8) is very intuitive. The sign of $R(e, w, \Gamma_j)$ is determined by the product $(w - 100\nu)(e - \mu)$. Suppose the worker gets a wage higher than 100ν ($100\nu\%$ of the maximum wage of 100), which the worker interprets as a kind offer. Then the worker responds by putting in an effort level higher than μ (i.e., a fraction μ of the maximum effort level of 1) to reciprocate the firm's kindness, as perceived by the worker. Analogously, unkind offers ($w < 100\nu$) by the firm are reciprocated by lower effort choices ($e < \mu$).

Substituting from (5.5), (5.6), (5.8) in (5.1), we get the *psychological utility* of a worker of type $j = I, N$ in the BGE model:

$$U(e,w,\Gamma_j) = w - 20e^2 + \lambda_R(w-100\nu)(e-\mu) - \lambda_G + \lambda_G e + \lambda_S \int_{x=0}^{x=e} P_W^2(x \mid \Gamma_j)\,dx + \lambda_G \int_{x=e}^{x=1} P_W^2(x \mid \Gamma_j)\,dx. \quad (5.9)$$

Remark 1 *(Beliefs in the competing models): In the CGE model (3.1), we had neither first order nor second order beliefs. In the AGE model (4.3), first order beliefs entered into the objective function through wage expectations (4.2) but no higher order beliefs were required. In the BGE model (5.9), we have only second order beliefs. These determine guilt-aversion (5.3), surprise-seeking (5.4), and intentions-driven reciprocity (see (5.8) and the proof of Proposition 3 where second order beliefs are used). If we find that the evidence rejects the CGE and the AGE models, but is consistent with the predictions of the BGE model, then we may conclude the following. The underlying mechanism that supports gift exchange is emotions such as guilt-aversion, surprise-seeking, and intentions-driven reciprocity.*

We now state our final assumption on beliefs, followed by a discussion.

Assumption A4 *We make the following assumption on the second order beliefs of the worker,* P_W^2

$$\frac{\partial P_W^2(e \mid \Gamma_j)}{\partial \gamma} < 0 \text{ for all } e \in (0,1),\ \gamma \in \Gamma_j,\ \gamma \in (0,\overline{\gamma}),\ j = I, N.$$

From Assumption A4 it follows that, for any $\gamma \in \Gamma_j$, $j = I, N$, a higher value of γ induces strict first order stochastic dominance in $P_W^2(e \mid \Gamma_j)$. This is formalized by Corollary 1, below.

Corollary 1 *(Strict first order stochastic dominance for second order beliefs of a worker): If* $0 \leq \gamma_1 < \gamma_2 \leq \overline{\gamma}$, *then* $P_W^2(e|\gamma_2,.) < P_W^2(e|\gamma_1,.)$ *for* $e \in (0,1)$.

Discussion of Assumption A4: $P_W^2(e \mid \Gamma_j)$ is the cumulative probability distribution of second order beliefs of the worker, which gives the worker's best guess of how much effort is expected by the firm. Assumption A4 requires that higher values of $\gamma \in \Gamma_j$ make it more likely to the worker that the firm expects a higher effort. For instance, if the firm operates in an industry where the typical effort, θ_e, is high, then the worker believes that it is more likely that the matched firm also expects the worker to put in a high effort level. Similarly, when the firm sends a high effort expectation signal, s, to the worker, then the worker believes it is more likely that the firm expects a high effort. Finally, in firms that operate in an industry where the typical wage, θ_w, is high, then the worker believes that the firm is more likely to expect a higher effort from the workers. In Section 12 we provided direct empirical tests of Assumption A4 that are strongly supportive.

Proposition 4 *The worker's utility (5.9) has a unique maximum,* $e^*(w,\Gamma_j)$. *Under Assumption A4 it has the following properties:*
(a) $\frac{\partial e^*(w,\Gamma_N)}{\partial s} = 0$.
(b) Suppose $w \leq 100\nu - \frac{\lambda_G}{\lambda_R}$. *Then* $e^*(w,\Gamma_j) = 0$.
(c) Suppose $100\nu - \frac{\lambda_G}{\lambda_R} < w < 100\nu + \frac{40-\lambda_S}{\lambda_R}$. *Then*

(ci) $\frac{\partial e^*(w,\Gamma_j)}{\partial w} > 0$,
(cii) $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta_w} > 0$, *for* $\theta_w \in (0,100)$,
(ciii) $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta_e} > 0$, *for* $\theta_e \in (0,1)$,
(civ) $\frac{\partial e^*(w,\Gamma_I)}{\partial s} > 0$, *for* $s \in (0,1)$.
(d) Suppose $w \geq 100\nu + \frac{40-\lambda_S}{\lambda_R}$. *Then* $e^*(w,\Gamma_j) = 1$.

Discussion of Proposition 4: At an interior solution, from (ci) the optimal effort levels of both informed and uninformed workers are increasing in the wage (gift exchange) as in the CGE and AGE models (Propositions 1, 2). The comparative statics with respect to θ_w, θ_e, s in Proposition 4 are different from those in the CGE and the AGE models (Propositions 1, 2) and are in accord with our evidence. An increase in θ_w, θ_e, s increases the probability, in the mind of the worker (only informed workers are influenced by s), that the firm expects higher effort from the worker (Assumption A4); and assumption A4 is directly empirically verified in Section 12. Hence, guilt-averse workers try to put in higher effort in order to reduce the possibility of ex-post guilt. Insofar as surprise-seeking is important, they also put in a higher effort to increase the possibility of ex-post elation.

Remark 2 *: From the discussion above, the positive effects of* θ_w *or* θ_e *on the optimal effort,* $e^*(w,\Gamma_j)$*, arise through the guilt-aversion channel. In particular, if we shut down this channel* ($\lambda_G = \lambda_S = 0$) *then we have* $\frac{\partial e^*(w,\Gamma_j)}{\partial \theta_w} = \frac{\partial e^*(w,\Gamma_j)}{\partial \theta_e} = 0$. *The reciprocity term in the objective function of the worker in (5.9) is* $\lambda_R(w - 100\nu)(e - \mu)$. *Recall that* μ *and* ν *are related, respectively, to the firm's and worker's entitlement to equitable payoffs,* as perceived by the worker. *The CGE and the AGE models do not predict sensitivity of the worker's effort to* μ, ν*, but the BGE model does. This could be an additional margin along which to distinguish between these models in future research, if empirical proxies for* μ, ν *could be found.*

6 Summary of the comparative static results

	$\frac{\partial e^*}{\partial w}$	$\frac{\partial e^*}{\partial \theta_w}$	$\frac{\partial e^*}{\partial \theta_e}$	$\frac{\partial e^*}{\partial s}$
Classical (CGE)	+	−	0	0
Augmented (AGE, informed worker)	+	−	−	−
Augmented (AGE, uninformed worker)	+	−	−	0
Belief-Based (BGE, informed worker)	+	+	+	+
Belief-Based (BGE, uninformed worker)	+	+	+	0

Table 3: A summary of the comparative static results under the different models when the optimal effort is an interior point.

Table 3 compares the predictions of the AGE and the CGE models (Propositions 1, 2) with the BGE model (Proposition 4) when $e^* \in (0,1)$. A "+" sign indicates that the corresponding derivative is positive; a "−" sign indicates a negative derivative; a "0" indicates that the derivative is zero. The BGE model makes different predictions to the other two models for each of the signals θ_w, θ_e, s. We can exploit these differences to empirically discriminate between the models.

7 Experimental Design

In subsection 7.1, below, we describe lab experiments designed to test the comparative static results for $\frac{\partial e^*}{\partial w}$, $\frac{\partial e^*}{\partial \theta_w}$, $\frac{\partial e^*}{\partial \theta_e}$ reported in Table 3, above. In subsection 7.2, we describe an online experiment we performed to test the comparative static results for $\frac{\partial e^*}{\partial s}$, also reported in Table 3. The restrictions around the coronavirus in both British and Chinese universities made it impossible to conduct lab experiments in the summer of 2020. Subsection 7.3 summarizes the treatments and sub-treatments described in subsections 7.1 and 7.2.

7.1 The lab experiments

The main experiments were conducted in March-May, 2019, in China with undergraduate and postgraduate university students from various disciplines. Subjects were randomly assigned to the roles of firms, informed workers, and uninformed workers. Their identities were kept anonymous from other players. Since we are mainly interested in the behavior of workers, assigning an equal number of subjects to firms and workers would have given us access to useful data for only half the subjects. For this reason, each firm was randomly matched with 4 workers. The workers who were matched with the same firm acted completely independently, i.e., they never observed each other's effort choices, nor did the choices of any one worker affect the material payoff of any other worker; the experimental instructions stressed and clarified this point. Each firm-worker interaction occurred only once. All this was common knowledge to the firms and workers. All four workers matched with a firm were offered the wage announced by the firm and along with their effort choice this determined their payoffs. The wage-effort choices when the firm was matched with one of the 4 workers, determined the payoffs of the firm. The wage offered by the firm and the effort chosen by any particular worker determined the payoff of the worker in that particular worker-firm interaction.

The experiments followed the sequence of moves that were described in detail in Section 2.2. Once the matching of firms and workers took place, the matched subjects were assigned the material payoffs given in (2.1) and (2.2). In order to ensure that the subjects fully understood the experimental instructions, we took the following steps (in addition to clear instructions). First, workers were provided with a table that showed the calculation of the cost of effort function $c(e) = 20e^2$ as effort increased from 0 to 1 in increments of 0.05.[21] Second, all subjects needed to successfully answer a series of control questions to proceed to the actual experiment.

The first 6 sessions (115 subjects) were conducted in Tianjin University of Finance and Economics, and the last 4 sessions (125 subjects) were conducted in Qingdao Agricultural University. There were 10 sessions in total involving 240 subjects; none of the subjects attended more than one session. There were 48 subjects in the role of the firm and 192 subjects in the role of the worker. Half the workers were randomly chosen as informed workers and the other half were chosen as uninformed workers. Thus, 96 subjects were assigned the role of informed workers, and 96 subjects were assigned the role of uninformed workers. There were three different treatments for each worker, and in each treatment a worker made 21 choices. Following the strategy method, we

[21] Subjects in the experiment could choose any effort level between 0 and 1.

collected 12,144 data points for the decisions made by firms and workers ($48 + 96 \times 2 \times 3 \times 21$). Each session lasted for around 1 hour. The currency used in the experiments was termed 'tokens', and subjects knew that the exchange rate was 1.5 tokens = 1 Yuan. Additionally, each subject received 20 Yuan as a participation fee for this experiment. On average, the subjects earned 48.45 Yuan. In May 2019, 1 Yuan was roughly equal to \$0.15. We ensured that subjects were never out of pocket.

7.1.1 Stage 1

In Stage 1, and as described in Section 2.2, the signals of (exogenous) wage and effort norms $\theta_w = 40$ and $\theta_e = 0.4$ were revealed to the firms.[22] Subjects were told that "the typical effort level exerted by workers in similar firms in the industry is $e = 0.4$" and "the typical wage paid by similar firms in the industry is $w = 40$ tokens." Then subjects in the role of firms were asked to guess the effort level they expected the matched worker to undertake. This 'guess' constitutes the signal s in our theoretical model. The responses of the firms were incentivized. If the difference between the firm's guess and the actual choice of effort by the matched worker was less than 0.1 (i.e., within an margin of error of 10%), then the firm earned an additional prize of 5 Yuan. The firm then chose a contractually binding wage, w.

7.1.2 Stage 2

Having observed the contractually binding wage, w, offered by the firm in Stage 1, the worker, in Stage 2, chooses a non-contractible effort, e. Stage 2 had two main treatments that differentiated between informed and uninformed workers. Each treatment had three sub-treatments (to study the comparative static effects of w, θ_w, θ_e). We describe these below.

1. **Treatment 1 (The uninformed worker)**: In order to test for the 3 comparative static effects with respect to w, θ_w, θ_e, we conducted three sub-treatments A, B, C; these are shown in columns 2, 3, 4, respectively, in Table 4. In each sub-treatment, we kept *fixed, two of these variables and varied the third*, using the strategy method to elicit the worker's effort response.[23] This effectively constituted a randomized controlled experiment that elicited the within-subject effort response of each worker when each of w, θ_w, θ_e was varied independently, keeping fixed the other two.

    - **Sub-treatment 1A (variable wage, w)**: Workers received the signals $\theta_w = 40$ and $\theta_e = 0.4$, which were identical to the signals conveyed to the firm in Stage 1. We then used the strategy method to determine the effort response $e \in [0, 1]$ of each worker to

[22] Our use of $\theta_w = 40$ and $\theta_e = 0.4$ was motivated by previous experimental studies of the gift exchange game which showed that the average wage level and the average effort level chosen by the subjects was around 40% of the maximum level; see e.g., Fehr et al. (1993), Fehr et al. (1998), Charness (2004), Charness and Kuhn (2007), Gächter et al. (2013).

[23] We did not randomize the order of sub-treatments A, B, and C for the following reason. Our workers were already split between the two groups of informed and uninformed workers. With 3 further comparative static effects, we would have needed 6 different randomizations of the order, leaving on average 16 workers for each sub-treatment. There would then be too few workers to accurately gauge order effects. Furthermore, and as reported below in Section 8.4, when we varied the order for one of the comparative static effects (with respect to θ_w and θ_e) there was no change in the results.

21 different equally spaced wage levels, w, from the interval $[0, 100]$ that could have been announced by the firm in Stage 1. This sub-treatment is designed to measure the comparative static effect, $\frac{\partial e^*}{\partial w}$.

- **Sub-treatment 1B (variable wage norm, θ_w)**: Workers were informed about the wage w set by the firm in Stage 1 and then the signal, $\theta_e = 0.4$ of the industry effort norm. Finally, using the strategy method, each worker was asked to choose an effort level, $e \in [0, 1]$ for each of 21 equally spaced hypothetical *wage norm* signals $\theta_w \in [0, 100]$. This sub-treatment is designed to measure the comparative static effect, $\frac{\partial e^*}{\partial \theta_w}$.
- **Sub-treatment 1C (variable effort norm, θ_e)**: Workers were informed about the wage w set by the firm in Stage 1 and the signal, $\theta_w = 40$, of the industry wage norm. We then used the strategy method to elicit the worker's optimal effort level, $e \in [0, 1]$, for each of 21 equally spaced hypothetical signals of *effort norms* $\theta_e \in [0, 1]$. This sub-treatment is designed to measure the comparative static effect, $\frac{\partial e^*}{\partial \theta_e}$.

  If any of the sub-treatments is played for real (after being randomly chosen) then one of the randomly chosen choices from the strategy method in that sub-treatment is implemented.

2. **Treatment 2 (The informed worker)**: Following the induced beliefs design (Ellingsen et al., 2010; Khalmetski et al., 2015; Dhami et al., 2019; Battigalli and Dufwenberg, 2022), in Treatment 2 (informed worker), the experimenter conveyed to the worker the effort level $s \in [0, 1]$ that the firm expects the worker to exert (elicited from the firm in Stage 1). Consistent with the induced beliefs design, workers knew that at the time the firm was asked to provide the signal s, in Stage 1, the firm was not told what the signal will be used for. Furthermore, the signal elicitation from the firm was incentive compatible (as described above) so that accurate (not strategic) guesses of the firm were rewarded. These precautions were taken to minimize concerns about strategic misrepresentation of the signal on the firm's part and to increase the confidence that workers had about the accuracy of the signal. None of our subjects in their roles as firms objected, ex-post, in the exit interviews, to their signals being conveyed to the workers and, in particular, none of them raised any concerns about subject deception. As with Treatment 1, in Treatment 2 we had three sub-treatments: 2A, 2B, 2C, where w, θ_w, θ_e were respectively varied.

7.2 The online experiment

We ran an extra subtreatment, subtreatment 2D, with informed workers. As in Balafoutas and Fornwagner (2017) we used the strategy method in which, for fixed levels of w, θ_w, θ_e, we elicited the effort response of the informed workers as we varied the signal s of the firm's expectations of the worker's effort between 0 to 1 in increments of 0.05. In this extra treatment, we chose the average wage in our earlier lab experiments ($w = 53$) and identical values $\theta_w = 40$ and $\theta_e = 0.4$ for the signals of wage and effort norms. Hence, we obtained exogenous variation in s, for each informed worker, that is independent of the wage, w, and can cleanly determine the relation between s and e for each worker. In order to ensure similarity of the subject pool with the main experiments,

we used undergraduate and postgraduate Chinese university students as the subjects of our online experiment.[24]

7.3 Summary of the treatments and sub-treatments

Treatment	w	θ_w	θ_e	s
1A	0 (5) 100	40	0.4	not given
1B	chosen by firm in Stage 1	0 (5) 100	0.4	not given
1C	chosen by firm in Stage 1	40	0 (0.05) 1	not given
2A	0 (5) 100	40	0.4	chosen by firm in Stage 1
2B	chosen by firm in Stage 1	0 (5) 100	0.4	chosen by firm in Stage 1
2C	chosen by firm in Stage 1	40	0 (0.05) 1	chosen by firm in Stage 1
2D	53	40	0.4	0 (0.05) 1

Table 4: Summary of the treatments. The notation $0(x)y$ means that the corresponding variable is varied from 0 to y in equal steps of x each.

For ease of reference, Table 4 summarizes (i) Treatment 1 and the three sub-treatments 1A, 1B, and 1C, and (ii) Treatment 2 and the four sub-treatments 2A, 2B, 2C and 2D. We use standard terminology in computational analysis, so that 0 (5) 100 means "varied from 0 to 100 in steps of 5" and 0 (0.05) 1 means "varied from 0 to 1 in steps of 0.05".

In addition, we collected information on demographic variables such as gender, education, and experience from the subjects.

8 Testing the comparative static results

In this section, we report the results of our empirical exercise. In sections 8.1 to 8.5 we examine the comparative static results with respect to the variables w, θ_w, θ_e, s summarized in columns 2-5 of Table 3. We first report the Spearman correlation coefficients between effort and the relevant variable for each individual worker, and then we report the histogram of regression coefficients across all workers, when effort is regressed sequentially on these variables; we have used a Tobit model censored at both tails in each case.

8.1 Relation between effort, e, and wage, w (gift exchange)

	Uninformed workers	Informed workers	Total
$\rho(e,w) > 0$	91(94.8%)	89(92.71%)	180(93.75%)
$\rho(e,w) = 0$	4(4.17%)	2(2.08%)	6(3.13%)
$\rho(e,w) < 0$	0(0%)	4(4.17%)	4(2.08%)

Table 5: Number and percent of the Spearman correlation coefficients between wage, w, and effort, e, which are significant at 1%.

In Table 5 we report the Spearman correlation coefficients between effort and wage, $\rho(e,w)$, that are significant at the 1% level, separately for informed and uninformed workers. For the

[24] We ran the online experiments using Wenjuanxing, which is a widely used Chinese platform providing functions that are equivalent to Amazon Mechanical Turk.

great majority, 93.75%, of the subjects, $\rho(e, w) > 0$, which is consistent with gift exchange; only 4 informed workers had negative coefficients. The data for the informed and uninformed workers shows similar levels of gift exchange. These results are consistent with all the three models, CGE, AGE, and BGE (see Propositions 1, 2, 4).

Figure 1 shows the distribution of the regression coefficients that are significant at 1% when the effort of each worker is regressed on the wage rate in a within-subjects design. We show the histograms for uninformed and informed workers in different colors. For ease of readability, the coefficients are multiplied by 100 on the horizontal axis. There is significant heterogeneity in the gift exchange behavior of the subjects, which is typical in such experiments.

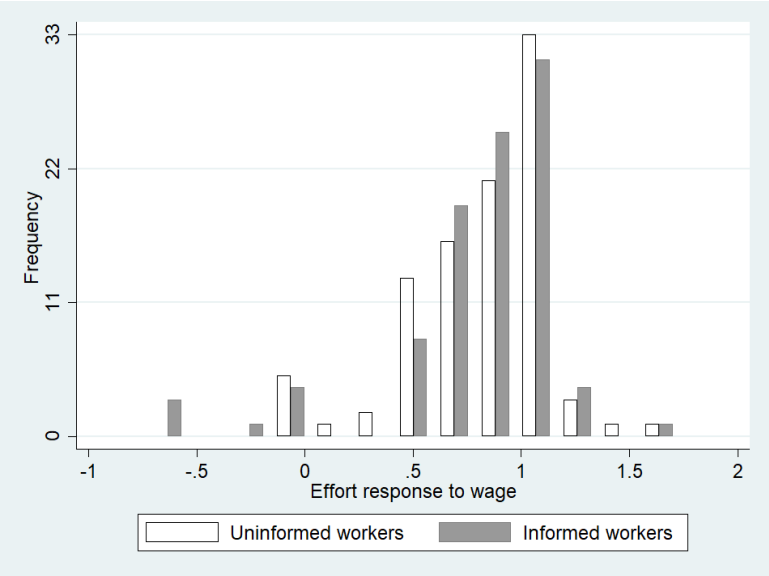

Figure 1: Histogram of regression coefficients, significant at 1%, when optimal effort, e, is regressed on wage, w, for informed and uninformed workers.

8.2 Relation between effort, e, and the signal of effort norm, θ_e

Table 6 shows the Spearman correlation coefficients between effort, e, and the signal of effort norm, θ_e, denoted by $\rho(e, \theta_e)$, that are significant at the 1% level. The great majority of the subjects exhibit $\rho(e, \theta_e) > 0$. From Table 3, the prediction of the CGE model is that $\frac{\partial e^*}{\partial \theta_e} = 0$; this prediction holds for 5.2% of the subjects, but it is rejected for the rest. The AGE model predicts that $\frac{\partial e^*}{\partial \theta_e} < 0$ which is true for only 1.04% of the subjects. By contrast, the prediction of the BGE model is that $\frac{\partial e^*}{\partial \theta_e} > 0$, and the behavior of 92.7% of the subjects is consistent with this prediction.

	Uninformed workers	Informed workers	Total
$\rho(e, \theta_e) > 0$	89(92.7%)	89(92.7%)	178(92.7%)
$\rho(e, \theta_e) = 0$	5(5.2%)	5(5.2%)	10(5.2%)
$\rho(e, \theta_e) < 0$	1(1.04%)	1(1.04%)	2(1.04%)

Table 6: Number and percent of the Spearman correlation coefficients between effort, e, and the typical effort in similar firms, θ_e, that are significant at 1%.

Figure 2 shows the histogram of the regression coefficients that are significant at 1% when we

regress effort on θ_e for the uninformed workers and the informed workers. Only two workers have significantly negative coefficients.

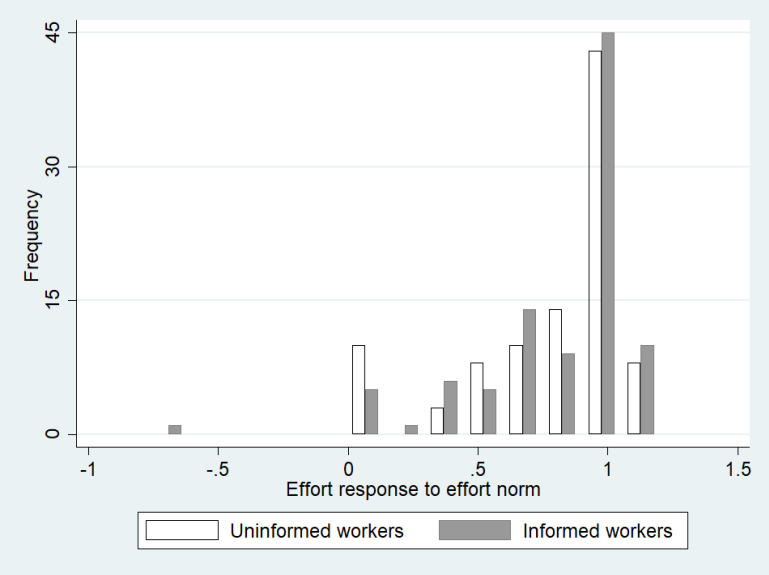

Figure 2: Histogram of regression coefficients significant at 1% when effort, e, is regressed on the effort level, θ_e, in similar firms, for informed and uninformed workers.

8.3 Relation between effort, e, and the signal of wage norm, θ_w

Table 7 shows the Spearman correlation coefficients between effort, e, and the signal of wage norm, θ_w, that are significant at the 1% level. The great majority of the subjects (68.8% across both types of workers) exhibit $\rho(e, \theta_w) > 0$, which is consistent with the prediction of the BGE model, $\frac{\partial e^*}{\partial \theta_w} > 0$ (Table 3). By contrast, the prediction of the CGE model and the AGE model is that $\frac{\partial e^*}{\partial \theta_w} < 0$, which holds for 17.7% of the subjects.

	Uninformed workers	Informed workers	Total
$\rho(e, \theta_w) > 0$	68(70.8%)	64(66.7%)	132(68.8%)
$\rho(e, \theta_w) = 0$	10(10.4%)	13(13.5%)	23(12.0%)
$\rho(e, \theta_w) < 0$	16(16.7%)	18(18.8%)	34(17.7%)

Table 7: Number and percent of the Spearman correlation coefficients between effort, e, and the typical wage, θ_w, in similar firms, that are significant at 1%.

Figure 3 shows the histogram of the regression coefficients that are significant at 1%, when we regress effort on θ_w for each worker. 13 informed workers and 0 uninformed workers have regression coefficients that are not significantly different from zero. 18 informed workers and 16 uninformed workers have significantly negative coefficients; their behavior is consistent with the CGE model. However, the vast majority of the informed and uninformed workers have significantly positive coefficients, which is consistent with the BGE model.[25]

[25]In Tables 5 and 6, one informed subject (1.04%) and one uninformed subject (1.04%) chose zero effort level at all levels of the independent variables, so their data was excluded. Similarly in Table 7, one informed worker (1.04%) and two uninformed workers (2.08%) chose zero effort level for all values of the independent variable, so their data was excluded as well. In the Figure 1, 2, 3, and 4, we did not exclude data. The subjects whose regression coefficient

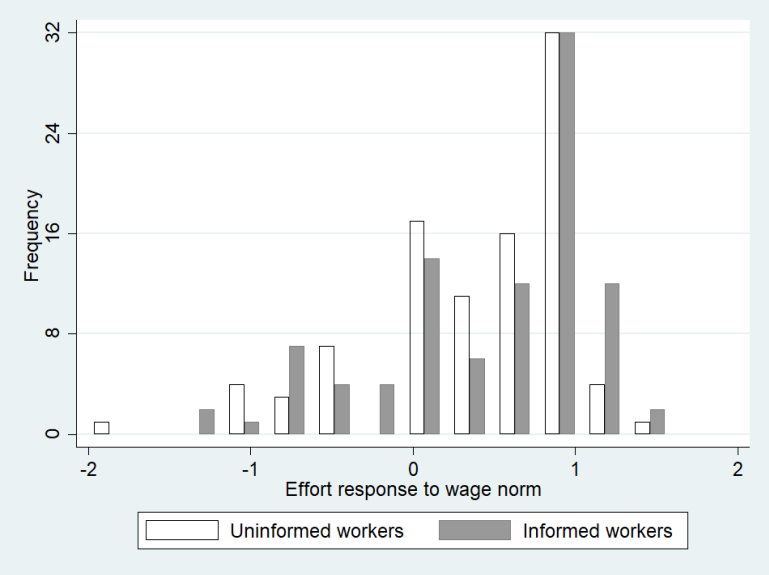

Figure 3: Histogram of regression coefficients significant at 1% when effort is regressed on the typical wage, θ_w, in similar firms for informed and uninformed workers.

8.4 Testing for fatigue and order effects

In order to test for the possibility that subjects might be fatigued with multiple calculations, we ran a separate session with 30 subjects. In this session, which corresponds to administering only sub-treatment B to the subjects (see Treatment 2B in Table 4), we only tested for the comparative static effect $\frac{\partial e^*}{\partial \theta_w}$. We do this by varying only θ_w between 0 and 100 in increments of 5 for 21 different values of θ_w, but we keep fixed the other two variables θ_e and w. For 71% of the subjects we still find $\rho\left(e, \theta_w\right) > 0$, which is very close to the figure of 68.8% in Table 7. This provides some assurance that (1) fatigue was not a factor in our results, and (2) there are no obvious concerns for order effects.

In order to test for the possibility that subjects might be fatigued with multiple calculations, we ran a separate session with 31 subjects. In this session, we administered only sub-treatment C to the subjects (see Treatment 2C in Table 4), so we only tested for the comparative static effects of the signal θ_e. We varied θ_e between 0 and 1 in increments of 0.05 for 21 different values of θ_e, keeping fixed everything else. For 87.1% of the subjects, we still found $\rho(e, \theta_e) > 0$, which is very close to the figure of 92.7% in Table 6. This provides more assurance that (1) fatigue was not a factor in our results, and (2) there are no obvious concerns for order effects.

8.5 Relation between effort, e, and the signal, s

In our extra sub-treatment, 2D, we ran an online experiment (see Section 7.2 for details). All 58 subjects in our online experiment were 'informed' workers because we were specifically interested in the effort response of informed workers when the signal s of the firm's effort expectations is exogenously varied. Subjects participated in the online experiment only once, and no subjects from the main experiment attended the online experiment. We dropped the data for 6 subjects,

is not significantly positive/negative at 1% level actually are categorized to the case that their regression coefficient is not significantly different from zero.

because they did not answer the control questions correctly, hence, we report the results for the remaining 52 subjects.[26]

The results were as follows. Among the 52 informed workers, 3 exerted zero effort at all levels of the signal, s. Since we used the strategy method to elicit each informed worker's best response for 21 different levels of the signal s, we can compute the Spearman correlation coefficient between effort and the signal, $\rho(e, s)$, for each worker. For 46 out of 52 informed workers, $\rho(e, s)$ is positive and significant at 1%. The average value of $\rho(e, s)$ across all 46 informed workers, whose Spearman correlation coefficient is significant, is 0.94. These results are consistent with the BGE model, but not with the CGE model and the AGE model (see column 5 in Table 3).

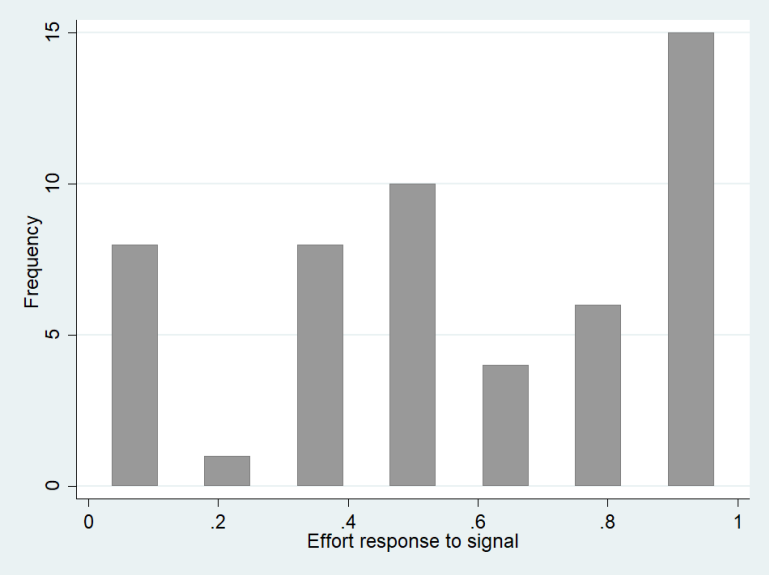

Figure 4: Histogram of regression coefficients, significant at 1%, when the optimal effort of informed workers is regressed on the signal, s.

Next, we ran Tobit regressions censored on both sides for each of the 52 informed workers. The effort choice is the dependent variable and the signal, s, is the independent variable. The distribution of the regression coefficients is shown in Figure 4. We found that 46 informed workers exhibited positive regression coefficients significant at 1%, and the regression coefficients of 3 informed workers were not significantly different from zero (recall that the remaining 3 informed workers chose zero effort for all values of the signal, s). The average regression coefficient of the 52 informed workers is 0.57. Excluding the two-direction changing cases where the effort level is non-monotonic in signals,[27] the average size of the regression coefficients is 0.6. The heterogeneity in responses across workers suggests possible heterogeneity in the parameter of guilt aversion across workers.

In sum, the statistical results strongly support the prediction of the BGE model but not the CGE model and the AGE model.

[26] The 52 subjects spent an average of 9 minutes completing the experiment, and received an average of around 14 Yuan.

[27] Among the 10 cases with non-monotonic effort responses to the signals, 8 of them exerted monotonically decreasing efforts at some value of the signals.

9 Determinants of effort

We now consider the determinants of effort choice in several Tobit models (censored on both sides) in Table 8, separately for informed workers (models 1–4) and uninformed workers (models 5–6). A discussion of the results and an evaluation of the alternative models follows in Section 10.

Table 8: Determinants of effort.

	Dependent Variable: Effort					
	Informed Workers				Uninformed Workers	
Tobit Model	1	2	3	4	5	6
Signal	0.399***	0.149**	0.367***	0.151*		
	[0.077]	[0.074]	[0.094]	[0.082]		
Wage		0.006***		0.005***	0.011***	0.011***
		[0.001]		[0.001]	[0.001]	[0.001]
Male			0.075**	0.051		-0.020
			[0.032]	[0.034]		[0.052]
Education			0.030***	0.019		0.018
			[0.012]	[0.012]		[0.019]
Experience			-0.019	-0.007		0.028
			[0.069]	[0.077]		[0.048]
Constant	0.232***	0.061	0.162***	0.035	-0.121**	-0.154
	[0.058]	[0.054]	[0.055]	[0.054]	[0.047]	[0.067]
Uncensored observations	95	95	95	95	95	95
F statistic	27.13***	34.97***	12.96***	35.26***	212.23***	60.17***
Log Likelihood	38.835	46.855	41.977	48.378	47.919	49.220

Note: Superscripts stars, ***, **, * denote significance levels of 1%, 5%, and 10%, respectively. Clustered standard errors are in brackets (clustering on experimental sessions).

The dependent variable is the effort choice of the worker. The independent variables are as follows: The signal received by the worker about the firm's first order belief, s (Signal); wage choice of the firm (Wage); dummy variable (Male) that equals 1 if the worker is male and 0 otherwise; dummy variable (Experience) that takes a value 1 if the subject has participated in similar experiments before and 0 otherwise; the variable 'Education' takes values from the set $\{1, 2, \ldots, 7\}$ and higher values denote higher educational attainment, e.g., Education = 1 for first year undergraduate students and Education = 6 for second year master students.[28]

In the experiments, the uninformed group did not receive the signal s of the firm's first order belief, and thus 'Signal' is not considered as an independent variable in Models 5 and 6 in Table 8. The wage norm, θ_w, and the effort norm, θ_e, cannot be included as part of the regression analysis because there is no variation in these variables in at least 2 out of the 3 sub-treatments (see a description of the sub-treatments in Section 7, and the summary in Table 4).

[28] In the post-experimental survey we also obtained data on the subjects' field of study, and aimed to investigate if there was difference in the choices between economics and non-economics students. However, there are no economics students in the uninformed group, and only 7 (out of 96) subjects are economics students in the informed group. Therefore this variable is omitted in the regressions in Table 8.

	$\frac{\partial e^*}{\partial w}$	$\frac{\partial e^*}{\partial \theta_w}$	$\frac{\partial e^*}{\partial \theta_e}$	$\frac{\partial e^*}{\partial s}$
Classical (CGE)	✓	×	×	×
Augmented (AGE)	✓	×	×	×
Belief-Based (BGE)	✓	✓	✓	✓

Table 9: A summary of the comparative static results under the different models. The last column applies to informed workers only.

For the informed workers, the explanatory variable Signal (s) is positive and significant. The positive sign is predicted for informed workers by the BGE model but not by the AGE and CGE models (see column 5 in Table 3). The effect of the wage is positive and significant in all regressions, for both types of workers, suggesting that gift exchange is an important feature of our data. This is consistent with all the models under consideration (see column 2 in Table 3). None of the personal and demographic variables are significant, once we include the wage, w, on the right hand side of the regression.

10 Discussion and summary of the results

Table 3 summarizes the comparative static results based on Propositions 1, 2, 4, when there is an interior solution to effort $e^* \in (0, 1)$. Following on, Table 9 evaluates the performance of the alternative models, based on the discussion in Sections 8 and 9. A "✓" denotes a confirmation and a "×" denotes a rejection of the relevant comparative static result by the data. For the vast majority of subjects, every prediction of the BGE model is confirmed by the data but three out of four predictions of the CGE and the AGE models are not consistent with the data.

From Tables 3 and 9 we see that all models explain well the basic prediction of gift exchange, $\frac{\partial e^*}{\partial w} > 0$. From Table 5 and Figure 1 we see that $\rho(e, w) > 0$ for 93.75% of the workers at the 1% level. However, only the BGE model can explain the observed positive effect on effort of increases in θ_w, θ_e and s.

Second order beliefs do not play any role in either the CGE or AGE models. By contrast, in the BGE model, second order beliefs play a crucial role, through the guilt-aversion channel, in deriving the three empirically correct results $\frac{\partial e^*}{\partial \theta_w} > 0$, $\frac{\partial e^*}{\partial \theta_e} > 0$ and (for the informed worker) $\frac{\partial e^*}{\partial s} > 0$. As mentioned in Remarks 1 and 2, this implies that the underlying mechanism that supports gift exchange is emotions such as guilt-aversion, surprise-seeking, and intentions-driven reciprocity. Thus, the evidence strongly supports the role of conditional second order beliefs in explaining gift exchange. Indeed, it is vital to derive testable predictions that are able to distinguish between the three models.

From Table 7 and Figure 3 we see that, at the 1% significance level, $\rho(e, \theta_w) > 0$ for 68.8% of the workers (and $\rho(e, \theta_w) < 0$ for 17.7% of the workers). This suggests that more than two-thirds of the workers indirectly engage in gift exchange with respect to the exogenous industry wage norm, θ_w, although they derive material utility only from the wage, w. This finding is also related to the literature that finds greater effort in the presence of minimum wage laws (Owens and Kagel, 2010), and gift exchange when the firm's intentionality cannot be determined (Charness, 2004;

Malmendier and Schmidt, 2017).

The finding of a positive effect of θ_w on effort is also consistent with Akerlof (1982), who uses a different transmission channel. In the Akerlof model, at a competitive general equilibrium, the wage norm in the industry also equals (i) the wage offered by the firm, and (ii) the reference wage of the worker. Thus, an increase in the wage norm is equivalent to an increase in the wage of the firm, which elicits gift exchange from the worker in terms of higher effort. Our contribution is to show theoretically, and verify empirically, a different transmission channel that arises from a rigorous formulation of the BGE model. We are not aware of stringent empirical evidence that tests the particular transmission channel in the Akerlof (1982) model, but it does not predict that informed workers should increase effort when they receive a higher signal s of the firm's effort expectation, which is a major finding of our paper.

11 Summary of the findings from the post-experimental survey

The results of the post-experimental, self-reported, non-incentivized, survey are as follows.

In making their effort choice, 83/96 uninformed workers and 82/96 informed workers report being influenced by θ_e (Q3 in the survey). Only 3.13% of the workers say that they decrease their effort when θ_e increases. Thus, the vast majority of workers increase their effort in response to an increase θ_e, which is consistent with the BGE model but not the other two models (see Table 3).

In choosing their effort, 81/96 uninformed workers and 80/96 informed workers report being influenced by θ_w (Q4 in the survey). Only 4.17% of the uniformed and 11.46% of the informed workers say that they decreased their effort when θ_w increases. Thus, a relatively high majority of the workers increases effort in response to an increase in θ_w, which is consistent with the BGE model but not the other two models (see Table 3).

In choosing their effort, 92/96 informed workers reported being influenced by the firm's expectation of effort, s (Q5 in the survey); this question was only relevant for informed workers. Out of these 92 workers, when the firm expects a higher effort, 14 informed workers always increased effort, 2 informed workers always decreased effort, 66 increase effort only if the wage, w, exceeds the typical wage, θ_w, in similar firms, and 10 informed workers decrease effort only if $w < \theta_w$. These results are inconsistent with the AGE and CGE models but consistent with the BGE model (see Table 3). They also suggest a more nuanced approach in which, at least for some workers, the guilt channel is operative in inducing higher effort, only if $w > \theta_w$. This empirical finding may guide future research.

12 Testing assumptions A3 and A4

In this section, we report on our testing of Assumptions A3 and A4. Our empirical tests strongly confirm both. What we do test is strict first order stochastic dominance for first and second order beliefs of a worker. As far as we are aware, this is the first direct empirical test of the assumptions of first order stochastic dominance of first and second order beliefs.

We show that strict first order stochastic dominance for first order beliefs of a worker implies Assumption A3. Hence, confirmation of the former is confirmation of Assumption A3. Assump-

tion A4 implies strict first order stochastic dominance for second order beliefs of a worker. We use Corollary 1 of Section 5, which is empirically indistinguishable from Assumption A4, to test Assumption A4.

12.1 Testing Assumption A3

To test Assumption A3 it is more convenient to test a stronger assumption which we introduce below, as Assumption A5. Proposition 5 then establishes that Assumption A5 implies Assumption A3.

Assumption A5 *The conditional first order beliefs of the worker satisfy:*

$$\frac{\partial P_W^1(w|\Gamma_j)}{\partial \gamma} < 0 \text{ for all } w \in (0,100),\ \gamma \in \Gamma_j,\ \gamma \in (0,\overline{\gamma}),\ j = I, N.$$

From Assumption A5 it follows that, for any $\gamma \in \Gamma_j$, $j = I, N$, a higher value of γ induces strict first order stochastic dominance in $P_W^1(w|\Gamma_j)$. This is formalized by Corollary 2 below.

Corollary 2 *(Strict first order stochastic dominance for first order beliefs of a worker): If* $0 \leq \gamma_1 < \gamma_2 \leq \overline{\gamma}$, *then* $P_W^1(w|\gamma_2, .) < P_W^1(w|\gamma_1, .)$ *for* $e \in (0,1)$.

Proposition 5 *: Assumption A5 implies Assumption A3.*

Following Corollary 2, which implies Assumptions A3 and A5, suppose $\gamma = \theta_w$, which is the wage offered in similar firms in the industry. Assume that we have two values of θ_w: A 'low value' $\theta_w = 40$ and a 'high value' $\theta_w = 80$.[29] We hold fixed the other two signals θ_e and s. Then Assumption A5 (and, hence, Assumption A3) holds if the worker believes that "it is more likely that the firm will offer a higher wage when $\theta_w = 80$ as compared to $\theta_w = 40$." Similar tests can be used for the cases $\gamma = \theta_e$ and $\gamma = s$. We employ this method below by varying one signal among θ_w, θ_e, s and keeping the other two fixed in three different between-subjects treatments.

12.2 Testing Assumption A4

Testing Assumption A4 requires us to check the second order beliefs of the worker about the first order beliefs of the firm about the worker's effort. Recall A4 requires us to check that:

$$\frac{\partial P_W^2(e \mid \Gamma_j)}{\partial \gamma} < 0 \text{ for all } e \in (0,1),\ \gamma \in \Gamma_j,\ \gamma \in (0,\overline{\gamma}),\ j = I, N.$$

Following Corollary 1, which implies Assumption A4, suppose once again that $\gamma = \theta_w$, and we have a 'low value' $\theta_w = 40$ and a 'high value' $\theta_w = 80$. We fix θ_e, the effort level that a typical worker exerts in a similar firm in the industry. Then, a direct test of Assumption A4 requires us to check that the worker believes that the firm is "more likely to expect a higher effort from the worker,

[29] Our choice of the low value of $\theta_w = 40$ is to keep it identical to the main experiment. The choice of high value of $\theta_w = 80$ is to distinguish it from the low value. Our tests of Assumptions A3 and A4 only require qualitative differences in responses on the likelihood of an event, and nothing hinges on the exact quantitative differences. A similar reasoning applies to our choice of low and high values of θ_e and s.

when $\theta_w = 80$ as compared to $\theta_w = 40$." This exercise can be repeated with the other signals and we follow this strategy below. Furthermore, we ensure incentive compatibility of responses and introduce additional questions for checking the consistency of responses.

12.3 The experimental design

We ran 3 online treatments to test A3 and A4 in the middle of March, 2022, at Nankai University with undergraduate and postgraduate students from various disciplines. Subjects were randomly assigned to the roles of firms and workers. Their identities were kept anonymous from other players. Each firm was matched with only one worker, and each firm-worker interaction occurred only once. All this was common knowledge to the firms and workers. In total, 182 subjects attended the experiment and no one attended more than one session. The experiment lasted around 11 minutes on average, and each subject earned around 28 Yuan[30]. To test A3 and A4, we varied only one signal at a time and kept the other signals fixed in each of 3 treatments, one treatment for each signal. Treatment 1 (T1) only varied θ_w; treatment 2 (T2) only varied θ_e, and treatment 3 (T3) only varied s.

Table 10: Description of the treatments

Treatment	Description	No. of subjects
T1	variation in θ_w only; Q1-Q5	62
T2	variation in θ_e only; Q1-Q5	60
T3	variation in s only; Q1-Q5	60

Table 10 shows a brief description and the number of subjects in each treatment. In each treatment, we asked 5 questions from the workers (Q1-Q5) that we describe below (details in the supplementary section).

In T1, θ_e is fixed at 0.4, s is fixed and elicited from the matched firm[31], but θ_w takes two values, $\theta_w = 40$ and $\theta_w = 80$. In T2, θ_w is fixed at 40, s is fixed and elicited from the matched firm, and θ_e takes two values $\theta_e = 0.4$ and $\theta_e = 0.8$. In T3, θ_e is fixed at 0.4, θ_w is fixed at 40, and s takes two values $s = 0.4$ and $s = 0.8$.

To test Assumption A5 (and, hence, Assumption A3), in each of the treatments, question Q3 asked the workers in which case ($\theta_w = 40$ or $\theta_w = 80$ in T1; $\theta_e = 0.4$ or $\theta_e = 0.8$ in T2; $s = 0.4$ or $s = 0.8$ in T3) they believe their matched firm is "more likely to offer a higher wage." Furthermore, in question Q4, the workers guessed the possible wage levels their matched firms would offer in the two cases. This was to ensure consistency of answers (between Q3 and Q4) and to ensure their incentive compatibility. If the subject's answer turned out to be within a margin of 10% of the target variable, then they could earn an additional prize of 30 tokens.

To test A4, for each treatment, question Q1 asked the workers in which case ($\theta_w = 40$ or $\theta_w = 80$ in T1; $\theta_e = 0.4$ or $\theta_e = 0.8$ in T2) they believe their matched firm is "more likely to expect a higher effort." This tests for the workers' second order beliefs. Furthermore, in question Q2, in order to check for the consistency of answers and to ensure incentive compatibility, the workers

[30] In March 2022, 1 Yuan was roughly equal to \$0.16.

[31] Although s might be different for different workers, it is fixed for each worker since it is elicited from their matched firm.

were also asked to state a numerical value of their guess of the possible effort expectation from their matched firms in the two cases. It would have been odd and redundant to ask the workers in which case they believe the matched firm expects a higher effort after revealing the matched firm's expectation of effort in the cases $s = 0.4$ or $s = 0.8$. Hence, and only in questions Q1 and Q2, we did not reveal the matched firm's actual effort expectation to the workers.

In the experiments, we placed questions Q1 and Q2 before questions Q3 and Q4. This ensured that the workers can legitimately guess the matched firm's expectation of their effort before they are revealed the matched firm's actual expectations of effort. For each treatment, we also asked in Q5, the effort that the worker would put in for further consistency checks and for incentive compatibility reasons (see below).

Prior to the 5 questions in each treatment, we also elicited information from the matched firm. For each of the treatments, we asked the firm the wage that it would offer and the effort expectation it held for the worker for each of the parameter configurations in Table 10. These effort expectations constituted the signal s that was provided to the worker. For instance, for Treatment 1, the firm was asked, for θ_e fixed at 0.4, what effort expectation, s, it has from the worker, when θ_w takes two values, $\theta_w = 40$ and $\theta_w = 80$. The firm was also asked what wage, w, it would offer in each of the two cases. This allowed us to tightly incentivize our procedure by (i) using the actual expectations of the firm in the worker's choice problem, and (ii) ensuring that we used actually chosen wage and effort levels to determine the material payoffs of firms and workers.

12.4 Results on testing Assumption A5 (and Assumption A3)

The empirical results strongly confirmed Assumption A5 and, hence, Assumption A3. For each of the three treatments, workers believed that the matched firm would offer a higher wage when the signal was high relative to when it was low (e.g., $\theta_w = 80$ is a higher signal relative to $\theta_w = 40$). In other words, we find strong support for Corollary 2 for our chosen parameter values.

Table 11: Percentage of subjects whose choices are consistent with Assumption A5. We also indicate the absolute number of subjects in each case (e.g., 25/31 means 25 subjects out of 31) and n is the number of subjects in each of the treatments in a between-subjects design.

Treatment	Q3	Q4	Q3 and Q4 consistent
T1 ($n = 31$)	25/31 (80.6%)	28/31 (90.3%)	23/24 (95.8%)
T2 ($n = 30$)	24/30 (80%)	23/30 (76.7%)	23/28 (82.1%)
T3 ($n = 30$)	24/30 (80%)	25/30 (83.3%)	24/27 (88.9%)

Table 11 summarizes the proportion of workers whose choices are consistent with Assumption A5. In treatment T1, 80.6% of the workers believed their matched firm is more likely to offer a higher wage when $\theta_w = 80$ relative to $\theta_w = 40$ in Q3; and 90.3% of the workers expected a higher wage in the case $\theta_w = 80$ relative to $\theta_w = 40$ in Q4. Some of the choices were found to be inconsistent. For example, some workers believed their matched firm is more likely to offer a higher wage when $\theta_w = 80$ in Q3, but expected a higher wage in the case $\theta_w = 40$ in Q4, which is a highly inconsistent response. After ruling out the inconsistent responses, the final column in Table 11 reports consistent choices; 95.8% of the workers' choices are consistent with Assumption

A5. Proceeding in this manner, the percentage of consistent choices that confirm Assumption A5 in the other treatments (T2 and T3) are, respectively, 82.1% and 88.9%.

In addition to the individual-level results above, the aggregate results are as follows. The workers' first order beliefs (FOBs) of their matched firm's wage offer, when each of the signals takes a high and a low value in the 3 treatments is as follows. In treatment T1, the average FOB when $\theta_w = 40$ is 44.92 tokens, and when $\theta_w = 80$ the average FOB 72.79 tokens (Mann-Whitney test, $p = 0.000$; one-sided t test, $p = 0.000$). In treatment T2, the average FOB in the case $\theta_e = 0.4$ is 39.13 tokens, while the average FOB in the case $\theta_e = 0.8$ is 59.12 tokens (Mann-Whitney test, $p = 0.007$; one-sided t test, $p = 0.033$). In treatment T3, the average FOB in the case $s = 0.4$ is 27.53 tokens, while the average FOB in the case $s = 0.8$ is 44 tokens (Mann-Whitney test, $p = 0.000$; one-sided t test, $p = 0.000$). We also conducted the non-parametric Mann-Whitney test and the parametric (one-sided) t test; both reveal very small p-values. Thus, we can conclude that the workers' FOBs are significantly higher under high values of the signals relative to low values of the signals, for each of the signals, which is consistent with Assumption A5 at the aggregate level.

12.5 Results on testing Assumption A4

The workers' choices are also strongly consistent with A4. In each of the three treatments, the workers believed that the matched firm expects a higher effort under a higher signal (e.g., $\theta_w = 80$ is a higher signal relative to $\theta_w = 40$). Hence, we find strong support for Corollary 1 for our chosen parameter values.

Table 12: Percentage of subjects whose choices are consistent with Assumption A4. We also indicate the absolute number of subjects in each case (e.g., 26/30 means 26 subjects out of 30) and n is the number of subjects in each of the treatments in a between-subjects design.

Treatment	Q1	Q2	Q1 and Q2 consistent
T1 ($n = 31$)	24/31 (77.4%)	27/31 (87.1%)	27/28 (96.4%)
T2 ($n = 30$)	23/30 (76.7%)	26/30 (86.7%)	22/24 (91.7%)

Table 12 summarizes the proportion of workers whose choices are consistent with Assumption A4. We do not have treatment T3 for testing A4; as noted above, it would be odd to ask subjects to guess the firm's expectation after revealing to them the firm's actual expectation. As in the case of Table 11, the last column in Table 12 shows choices that are consistent with each other in the sense that the answers to the different questions are not in conflict.

In treatment T1, 77.4% of the workers believed their matched firm is "more likely to expect a higher effort" in the case $\theta_w = 80$ relative to $\theta_w = 40$ in Q1. For Q2 that directly asked workers to put a numerical value on the effort they believed the firm expected of them, 87.1% believed the matched firm would expect a higher effort in the case $\theta_w = 80$ relative to $\theta_w = 40$. After ruling out the workers whose answers in Q1 and Q2 are not consistent[32], 96.4% of the workers' choices are consistent with A4. Similarly, for treatment T2, while restricting attention to consistent choices, 91.7% of the choices are consistent with A4.

[32]For example, some workers believed their matched firm is more likely to expect a higher effort in the case $\theta_w = 80$ in Q1, but their stated numerical effort that the firm expected of them was higher for $\theta_w = 40$ in Q2. Such choices are inconsistent.

In addition to the individual-level results above, the aggregate results are as follows. We compared the workers' beliefs of their matched firm's expectation of effort, which are the workers' second order beliefs (SOBs), under the low and the high signals in each treatment. In T1, the average SOB in the case $\theta_w = 40$ is 0.4, while the average SOB in the case $\theta_w = 80$ is 0.67 (Mann-Whitney test, $p = 0.000$; one-sided t test, $p = 0.000$). In treatment T2, the average SOB in the case $\theta_e = 0.4$ is 0.43, while the average SOB in the case $\theta_e = 0.8$ is 0.75 (Mann-Whitney test, $p = 0.000$; one-sided t test, $p = 0.000$). Furthermore, the non-parametric Mann-Whitney test and the parametric (one-sided) t test reveal very small p-values. Consequently, we can conclude the worker's SOBs are significantly higher under the high signals as compared to the low signals, which is consistent with A4 at the aggregate level.

12.6 Further consistency checks

In our main experiments, since we were mainly interested in data from workers, we had matched 1 firm with 4 workers. All workers operated independently of each other. This was well understood by our subjects as evidenced by their responses to the test of understanding of the instructions, which we necessary to proceed with the experiment (see Section 7 for the other safeguards that we used). In the newer experiments, we matched one worker with one firm, but the results are unchanged. This should remove any residual doubts about our earlier procedure. In the new treatments T1, T2, and T3, question Q5 asked workers to make their effort choices under different parameteric configurations in Table 10. In treatment T1, workers exerted higher effort in the case $\theta_w = 80$ than $\theta_w = 40$ (Mann-Whitney test, $p = 0.000$; one-sided t test, $p = 0.001$). In treatment T2, workers exerted higher effort in the case $\theta_e = 0.8$ than $\theta_e = 0.4$ (Mann-Whitney test, $p = 0.001$; one-sided t test, $p = 0.004$). In treatment T3, workers exerted higher effort in the case $s = 0.8$ than $s = 0.4$ (Mann-Whitney test, $p = 0.038$; one-sided t test, $p = 0.025$). The results on effort choices are consistent with the results of our main experiment.

13 Summary and conclusion

We considered three models of gift exchange, the classical (CGE), the augmented (AGE) and the belief-based (BGE) model. The BGE model explicitly uses belief hierarchies (beliefs and beliefs about beliefs) and is based on psychological game theory. First order (but not second order) beliefs enter the AGE model through the workers' expectations of the wage rate. Second order (but not first order) beliefs enter the BGE model through guilt aversion.

All three models are able to explain the gift-exchange hypothesis, namely, higher wages induce higher effort. However, only the BGE model is able to explain the optimal effort response to a change in the three signals in our model– a signal of the typical wage level in similar firms θ_w (or 'wage norm' in Akerlof's terminology); a signal of the typical effort level in similar firms θ_e (or 'effort norm' in Akerlof's terminology); and a signal of the firm's expectation of the effort of workers, s. By contrast, the CGE and the AGE models fail to make the correct prediction for the first two signals, θ_w, θ_e, and make no prediction for the third signal, s.

The main operative channel in the BGE model is through second order beliefs (beliefs of the

worker about the beliefs of the firm). This allows a formal and rigorous modeling of *guilt* and *conditional reciprocity* that underpin the comparative static results in the BGE model. It also allows us to derive predictions on the different implications of the three competing models. Our paper locates the microfoundations of gift exchange in second order, rather than first order, beliefs and clarifies the relevant emotions that underpin gift exchange. In this sense, our paper may also be seen to contribute to a more general emerging view in behavioral economics of the relative importance of explicitly modeling the underlying belief hierarchies.

Appendix: Proofs

Proof of Proposition 1: From (3.1) we see that W is a twice continuously differentiable function of $e \in [0,1]$. Furthermore, we get

$$\frac{\partial W\left(e;w,\Gamma_j\right)}{\partial e} = -40e + \beta\left(w-\theta_w\right), \tag{13.1}$$

$$\frac{\partial^2 W\left(e;w,\Gamma_j\right)}{\partial e^2} = -40 < 0. \tag{13.2}$$

From (13.2), we see that W is a strictly concave function of $e \in [0,1]$. Hence, a unique maximizer, $e^*\left(w,\Gamma_j\right) \in [0,1]$, exists. If $w \leq \theta_w$ then, from (13.1), we get $\left[\frac{\partial W(e;w,\Gamma_j)}{\partial e}\right]_{e=0} \leq 0$. Since W is strictly concave, we get $\frac{\partial W(e;w,\Gamma_j)}{\partial e} < 0$, for all $e \in (0,1]$. Hence, $e^*\left(w,\Gamma_j\right) = 0$. This establishes part (a). If $w \geq \frac{40}{\beta} + \theta_w$ then, from (13.1), we get $\left[\frac{\partial W(e;w,\Gamma_j)}{\partial e}\right]_{e=1} \geq 0$. Since W is strictly concave, we get $\frac{\partial W(e;w,\Gamma_j)}{\partial e} > 0$, for all $e \in [0,1)$. Hence, $e^*\left(w,\Gamma_j\right) = 1$. This establishes part (c).

Now suppose that $\theta_w < w < \frac{40}{\beta} + \theta_w$. From (13.1) we then get $\left[\frac{\partial W(e;w,\Gamma_j)}{\partial e}\right]_{e=0} > 0$ and $\left[\frac{\partial W(e;w,\Gamma_j)}{\partial e}\right]_{e=1} < 0$. It follows that $e^*\left(w,\Gamma_j\right) \in (0,1)$. Consequently, $\left[\frac{\partial W(e;w,\Gamma_j)}{\partial e}\right]_{e=e^*(w,\Gamma_j)} = 0$. Solving this equation, using (13.1), we get $e^*\left(w,\Gamma_j\right) = \frac{\beta}{40}\left(w-\theta_w\right)$. Part (b) then follows. ■

Proof of Proposition 2: From (4.3) we see that V is a twice continuously differentiable function of $e \in [0,1]$. Furthermore, we get

$$\frac{\partial V\left(e;w,\Gamma_j\right)}{\partial e} = -40e + \sigma\left[w - E\left(w \mid \Gamma_j\right)\right], \tag{13.3}$$

$$\frac{\partial^2 V\left(e;w,\Gamma_j\right)}{\partial e^2} = -40 < 0. \tag{13.4}$$

From (13.4), we see that V is a strictly concave function of $e \in [0,1]$. Hence, a unique maximizer, $e^*\left(w,\Gamma_j\right) \in [0,1]$, exists.

If $w \leq E\left(w \mid \Gamma_j\right)$ then, from (13.3), we get $\left[\frac{\partial V(e;w,\Gamma_j)}{\partial e}\right]_{e=0} = \sigma\left[w - E\left(w \mid \Gamma_j\right)\right] \leq 0$. Since V is strictly concave, we get $\frac{\partial V(e;w,\Gamma_j)}{\partial e} < 0$, for all $e \in (0,1]$. Hence, $e^*\left(w,\Gamma_j\right) = 0$ for $w \leq E\left(w \mid \Gamma_j\right)$. This establishes part (a).

If $w \geq \frac{40}{\sigma} + E\left(w \mid \Gamma_j\right)$ then, from (13.3), we get $\left[\frac{\partial V(e;w,\Gamma_j)}{\partial e}\right]_{e=1} = -40 + \sigma\left[w - E\left(w \mid \Gamma_j\right)\right] = \sigma\left\{w - \left[\frac{40}{\sigma} + E\left(w|\Gamma_j\right)\right]\right\} \geq 0$. Since V is strictly concave, we get $\frac{\partial V(e;w,\Gamma_j)}{\partial e} > 0$, for all $e \in [0,1)$. Hence, $e^*\left(w,\Gamma_j\right) = 1$ for $w \geq \frac{40}{\sigma} + E\left(w \mid \Gamma_j\right)$. This establishes part (c).

Now suppose that $E(w \mid \Gamma_j) < w < \frac{40}{\sigma} + E(w \mid \Gamma_j)$. From (13.3) we then get $\left[\frac{\partial V(e;w,\Gamma_j)}{\partial e}\right]_{e=0} = \sigma[w - E(w \mid \Gamma_j)] > 0$ and $\left[\frac{\partial V(e;w,\Gamma_j)}{\partial e}\right]_{e=1} = \sigma\left\{w - \left[\frac{40}{\sigma} + E(w|\Gamma_j)\right]\right\} < 0$. It follows that $e^*(w,\Gamma_j) \in (0,1)$. Consequently, $\left[\frac{\partial V(e;w,\Gamma_j)}{\partial e}\right]_{e=e^*(w,\Gamma_j)} = 0$. Solving this equation, using (13.3), gives

$$e^*(w,\Gamma_j) = \frac{\sigma}{40}[w - E(w \mid \Gamma_j)]. \tag{13.5}$$

Parts (bi)-(biv) then follow by direct differentiation of (13.5) with respect to $\gamma \in \Gamma_j$ and using Assumption A3. ■

Proof of Proposition 3: We give the proof in three parts, a, b and c.

(a) $k_{WF}(e,w) = 100(e-\mu)$.

From (2.1) we get that $\max\{\pi(e,w), e \in [0,1]\} = 100 - w + \kappa$ and $\min\{\pi(e,w), e \in [0,1]\} = -w + \kappa$. Hence, from Definition 2a, we get $\pi^E(w) = 100\mu - \mu w + \mu\kappa - w + \mu w + \kappa - \mu\kappa = 100\mu - w + \kappa$. From this, (2.1) and Definition 2b, we get $k_{WF}(e,w) = 100e - w + \kappa - (100\mu - w + \kappa) = 100(e-\mu)$.

(b) $\widehat{k}_{FW}(w,\Gamma_j) = w - 100\nu$.

From (2.2) and Definition 2c we get $Eu(w,\Gamma_j) = w - 20\int_{e=0}^{e=1} e^2 dP_W^2(e \mid \Gamma_j)$, $j = I, N$. Hence, $\max\{Eu(w,\Gamma_j), w \in [0,100]\} = 100 - 20\int_{e=0}^{e=1} e^2 dP_W^2(e \mid \Gamma_j)$, $j = I, N$ and $\min\{Eu(w,\Gamma_j), w \in [0,100]\} = -20\int_{e=0}^{e=1} e^2 dP_W^2(e \mid \Gamma_j)$, $j = I, N$. From these, and Definition 2d, we get $u^E(w,\Gamma_j) = 100\nu - 20\nu\int_{e=0}^{e=1} e^2 dP_W^2(e|\Gamma_j) - 20(1-\nu)\int_{e=0}^{e=1} e^2 dP_W^2(e \mid \Gamma_j)$. From Definition 2e it then follows that $\widehat{k}_{FW}(w,\Gamma_j) = w - 100\nu$.

(c) From Definition 2f and parts (a) and (b) it follows that $R(e,w,\Gamma_j) = 100(e-\mu)(w - 100\nu)$. ■

Proof of Corollary 1: This is an immediate consequence of Assumption A4 and the definition of first order stochastic dominance. ■

Proof of Proposition 4: From (5.9) we see that U is a twice continuously differentiable function of $e \in [0,1]$. Furthermore, we get

$$\begin{aligned}
\frac{\partial}{\partial e}U(e,w,\Gamma_j) &= -40e + \lambda_R(w - 100\nu) + \lambda_G \\
&\quad -(\lambda_G - \lambda_S)P_W^2(e|\Gamma_j), && (13.6)\\
\frac{\partial^2}{\partial e^2}U(e,w,\Gamma_j) &= -40 - (\lambda_G - \lambda_S)\frac{\partial}{\partial e}P_W^2(e \mid \Gamma_j), && (13.7)\\
\frac{\partial^2}{\partial e \partial w}U(e,w,\Gamma_j) &= \lambda_R - (\lambda_G - \lambda_S)\frac{\partial}{\partial w}P_W^2(e \mid \Gamma_j) = \lambda_R, && (13.8)\\
\frac{\partial^2}{\partial e \partial \theta_w}U(e,w,\Gamma_j) &= -(\lambda_G - \lambda_S)\frac{\partial}{\partial \theta_w}P_W^2(e \mid \Gamma_j), && (13.9)\\
\frac{\partial^2}{\partial e \partial \theta_e}U(e,w,\Gamma_j) &= -(\lambda_G - \lambda_S)\frac{\partial}{\partial \theta_e}P_W^2(e \mid \Gamma_j), && (13.10)\\
\frac{\partial^2}{\partial e \partial s}U(e,w,\Gamma_I) &= -(\lambda_G - \lambda_S)\frac{\partial}{\partial s}P_W^2(e|\Gamma_I). && (13.11)
\end{aligned}$$

From (5.2) and (13.7), since $\frac{\partial}{\partial e}P_W^2(e \mid \Gamma_j) \geq 0$, we get that $\frac{\partial^2}{\partial e^2}U(e,w,\Gamma_j) \leq -40 < 0$, $j = I, N$. Hence, $U(e,w,\Gamma_j)$ is a continuous and strictly concave function of $e \in [0,1]$. Therefore, a unique maximizer $e^*(w,\Gamma_j)$ exists.

From (5.9) and Definition 1 we see that $U(e; w, \Gamma_N)$ does not depend on s. Hence, neither can $e^*(w, \Gamma_N)$. This establishes part (a).

If $w \leq 100\nu - \frac{\lambda_G}{\lambda_R}$ then, from (5.2) and (13.6), we get $\left[\frac{\partial U(e;w,\Gamma_j)}{\partial e}\right]_{e=0} \leq 0$. Since U is strictly concave, we get $\frac{\partial U(e;w,\Gamma_j)}{\partial e} < 0$, for all $e \in (0,1]$. Hence, $e^*(w, \Gamma_j) = 0$. This establishes part (b).

If $w \geq 100\nu + \frac{40-\lambda_S}{\lambda_R}$ then, from (5.2) and (13.6), we get $\left[\frac{\partial U(e;w,\Gamma_j)}{\partial e}\right]_{e=1} \geq 0$. Since U is strictly concave, we get $\frac{\partial U(e;w,\Gamma_j)}{\partial e} > 0$, for all $e \in [0,1)$. Hence, $e^*(w, \Gamma_j) = 1$. This establishes part (d).

Now suppose that $100\nu - \frac{\lambda_G}{\lambda_R} < w < 100\nu + \frac{40-\lambda_S}{\lambda_R}$. From (5.2) and (13.6) we then get $\left[\frac{\partial U(e;w,\Gamma_j)}{\partial e}\right]_{e=0} > 0$ and $\left[\frac{\partial U(e;w,\Gamma_j)}{\partial e}\right]_{e=1} < 0$. It follows that $e^*(w, \Gamma_j) \in (0,1)$. Consequently,

$$\frac{\partial e^*(w,\Gamma_j)}{\partial w} = -\left[\frac{\frac{\partial^2}{\partial e \partial w}U(e,w,\Gamma_j)}{\frac{\partial^2}{\partial e^2}U(e,w,\Gamma_j)}\right]_{e=e^*(w,\Gamma_j)}, \tag{13.12}$$

$$\frac{\partial e^*(w,\Gamma_j)}{\partial \gamma} = -\left[\frac{\frac{\partial^2}{\partial e \partial \gamma}U(e,w,\Gamma_j)}{\frac{\partial^2}{\partial e^2}U(e,w,\Gamma_j)}\right]_{e=e^*(w,\Gamma_j)}, \tag{13.13}$$

$$\gamma \in \Gamma_j,\ J \in \{I, N\}. \tag{13.14}$$

Part (c) then follows from (5.2), (13.7)-(13.14), Definition 1 and Assumption A4. ■

Proof of Proposition 5: Integrating 4.2 by parts, we get: $E(w|\Gamma_j) = 100 - \int_{w=0}^{100} P_W^1(w|\Gamma_j)\, dw$ and, hence, $\frac{\partial}{\partial \gamma}E(w|\Gamma_j) = -\int_{w=0}^{100} \frac{\partial}{\partial \gamma}P_W^1(w|\Gamma_j)\, dw > 0$, for all $\gamma \in \Gamma_j$, $\gamma \in (0, \overline{\gamma})$, $j = I, N$. It follows that, if Assumption A5 holds, then Assumption A3 also holds. ■

14 Acknowledgements

We are grateful to Martin Dufwenberg, Junaid Arshad, Yan Chen, and Stephanie W. Wang for comments and suggestions to this and earlier versions of the paper. Some of the basic ideas were presented in seminars in LMU, Munich, The Kiel Institute for the World Economy, 2020 ESA Global Virtual Conference, Indian Statistical Institute, Delhi, and the Universities of Cardiff and Birmingham. We would also like to acknowledge funding from the following sources. National Natural Science Foundation of China, 72003100; Fellowship of China Postdoctoral Science Foundation, 2020M670616; and Fundamental Research Funds for the Central Universities, 63202022. Finally we are very grateful to the editor, the associated editor, and two referees for their comments and for helping to improve the paper. Declarations of interest: none.

References

[1] Abeler, J., Altmann, S., Kube, S., and Wibral, M. (2010). Gift exchange and workers' fairness concerns: when equality is unfair. Journal of the European Economic Association 8(6): 1299–324.

[2] Akerlof, George A. (1982). Labor Contracts as Partial Gift Exchange. The Quarterly Journal of Economics 97 (4): 543–69.

[3] Attanasi, G., Rimbaud, C., Villeval, M. C. (2019) Embezzlement and guilt aversion. Journal of Economic Behavior & Organization 167: 409-429.

[4] Balafoutas, L., and Fornwagner, H. (2017). The Limits of Guilt. Journal of the Economic Science Association 3: 137-148.

[5] Battigalli, P., Corarro, R. and Dufwenberg, M. (2019) Incorporating belief-dependent motivation in games. Journal of Economic Behavior and Organization 167: 185–218.

[6] Battigalli P., Dufwenberg M. (2007) Guilt in games. The American economic review. 97(2): 170-176.

[7] Battigalli P., Dufwenberg M. (2009) Dynamic psychological games. Journal of Economic Theory. 144(1): 1-35.

[8] Battigalli P., Dufwenberg M. (2022) Belief-Dependent Motivations and Psychological Game Theory. Forthcoming Journal of Economic Literature.

[9] Bejarano, H., Corgnet, B., and Gomez-Minambres, J. (2021). Economic stability promotes gift-exchange in the workplace. Journal of Economic Behavior & Organization 187: 374-398.

[10] Bellemare, C. and Shearer, B. (2009). Gift Giving and Worker Productivity: Evidence from a Firm Level Experiment. Games and Economic Behavior 67: 233-244.

[11] Bellemare, C., Sebald, A. and Strobel, M. (2011). Measuring the Willingness to Pay to Avoid Guilt: Estimation Using Equilibrium and Stated Belief Models. *Journal of Applied Econometrics* 26(3): 437–453.

[12] Bellemare, C., Sebald, A. and Suetens, S. (2019). Guilt aversion in economics and psychology. Journal of Economic Psychology 73: 52-59.

[13] Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge University Press, New York, USA.

[14] Blau, P. (1964). Exchange and Power in Social Life. New York: Wiley.

[15] Bowles, S. and Gintis, H. (2011). A cooperative species: Human reciprocity and its evolution. Princeton University Press: Princeton.

[16] Camerer C. (2003) Behavioral game theory: Experiments in strategic interaction. Princeton University Press: Princeton.

[17] Cartwright, E. (2019). A Survey of Belief-Based Guilt Aversion in Trust and Dictator Games. Journal of Economic Behavior and Organization 167: 430-444.

[18] Charness, G. (2004). Attribution and reciprocity in a simulated labor market: An experimental investigation. Journal of Labor Economics. 22(3), 665–688.

[19] Charness and Kuhn (2007). Does pay inequality affect worker effort? Experimental evidence. Journal of Labor Economics, 25(4), 693–723.

[20] DellaVigna, S., List, J. A., and Malmendier, U., and Rao, G. (2020) Estimating Social Preferences and Gift Exchange with a Piece-Rate Design. CEPR Discussion Paper No. DP14931, Available at SSRN: https://ssrn.com/abstract=3638035.

[21] Dhami, S. (2016) Foundations of behavioral economic analysis. Oxford University Press: Oxford.

[22] Dhami, S. (2019) Foundations of behavioral economic analysis: Volume II: Other-Regarding preferences. Oxford University Press: Oxford.

[23] Dhami, S. (2020) Foundations of behavioral economic analysis: Volume IV: Behavioral Game Theory. Oxford University Press: Oxford.

[24] Dhami, S. (2020) Foundations of behavioral economic analysis: Volume V: Bounded Rationality. Oxford University Press: Oxford.

[25] Dhami, S., Wei, M., and al-Nowaihi, A. (2019). Public goods games and psychological utility: Theory and evidence. Journal of Economic Behavior & Organization. 167: 361–390.

[26] Dhami, S., Arshad, J. and al-Nowaihi, A. (2020). Psychological and Social Motivations in Microfinance Contracts: Theory and Evidence (June 2020). CESifo Working Paper No. 7773, Available at SSRN: https://ssrn.com/abstract=3432821.

[27] Di Bartolomeo, G., Dufwenberg, M., Papaa, S., and Passarelli, F. (2018) Promises, expectations & causation. Games and Economic Behavior 113: 136-146.

[28] Dufwenberg M., Gächter S., Hennig-Schmidt H. (2011) The framing of games and the psychology of play. Games and Economic Behavior. 73(2): 459-478.

[29] Dufwenberg, M., and Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. Games and Economic Behavior. 47(2): 268–98.

[30] Dufwenberg, M., and Kirchsteiger, G. (2019). Modelling kindness. Journal of Economic Behavior & Organization. 167: 228–234.

[31] Ellingsen, T., Johannesson, M., Tjøtta, S. and Torsvik, G. (2010). Testing Guilt Aversion. *Games and Economic Behavior* 68(1): 95–107.

[32] Elster, J. (2011). Norms, in P. Bearman and P. Hedström, eds., The Oxford Handbook of Analytical Sociology. Oxford University Press, Oxford, UK, pp. 195–217.

[33] Englmaier, F., and Leider, S. (2012). Contractual and organizational structure with reciprocal agents. American Economic Journal: Microeconomics. 4(2): 146–183.

[34] Falk, A. (2007). Gift exchange in the field. Econometrica. 75(5): 1501–1511.

[35] Falk, A., and Fischbacher, U. (2006). A Theory of reciprocity. Games and Economic Behavior. 54(2): 293–315.

[36] Fehr, E., Gächter, S., and Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: experimental evidence. Econometrica. 65(4): 833–860.

[37] Fehr, Ernst, Goette, L. and Zehnder, C. (2009). A Behavioral Account of the Labor Market: The Role of Fairness Concerns. Annual Review of Economics, 1 (1): 355–84.

[38] Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. Quarterly Journal of Economics. 108(2): 437–459.

[39] Fehr, E., Kirchsteiger, G., and Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. European Economic Review. 42(1): 1–34.

[40] Fehr, E., and Schmidt, K. M. (1999). A Theory of fairness, competition, and cooperation. The Quarterly Journal of Economics. 114 (3): 817–68.

[41] Fehr, E., Fischbacher, U., and Tougareva, E. (2002). Do high stakes and competition remove reciprocal fairness? Evidence from Russia. University of Zurich, Institute for Empirical Research in Economics, Working paper 120.

[42] Fehr, E., and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism - experimental evidence and new theories. Handbook of the Economics of Giving, Altruism and Reciprocity. 1, 615–691.

[43] Gächter, S., Nosenzo, D., and Sefton, M. (2012). The impact of social comparisons on reciprocity. Scandinavian Journal of Economics, 114(4): 1346–67.

[44] Gächter, S., Nosenzo, D., and Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences?. Journal of the European Economic Association, 11(3), 548–573.

[45] Gächter, S. and Thöni, C. (2010): Social Comparison and Performance: Experimental Evidence on the Fair Wage-Effort Hypothesis. Journal of Economic Behavior & Organization, 76(3): 531-543.

[46] Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. Games and Economic Behavior. 1(1): 60–79.

[47] Gneezy, U., and List, J.A. (2006). Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. Econometrica. 74(5): 1365–1384.

[48] Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review* 25(2): 161–78.

[49] Hauge, K. A. (2016) Generosity and guilt: The role of beliefs and moral standards of others. Journal of Economic Psychology 54: 35-43.

[50] Henrich, J. (2016). *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*, Princeton University Press, NJ, USA.

[51] Inderst, R., Khalmetski, K., and Ockenfels, A. (2019). Sharing Guilt: How Better Access to Information May Backfire. Management Science 65(7): 2947-3448.

[52] Kandel, E., and Lazear, E. P. (1992). Peer pressure and partnerships. Journal of Political Economy. 100(4): 801-817.

[53] Khalmetski K., Ockenfels A, Werner P. (2015) Surprising gifts: Theory and laboratory evidence. Journal of Economic Theory. 159: 163–208.

[54] Kube, S., Maréchal, M. A., and Puppe, C. (2012). The Currency of Reciprocity: Gift Exchange in the Workplace. American Economic Review 102(4): 1644-1662.

[55] Kube, S., Maréchal, M. A., and Puppe, C. (2013). Do Wage Cuts Damage Work Morale? Evidence From A Natural Field Experiment. Journal of the European Economic Association 11(4): 853-870.

[56] Levine, D. K. (1998) Modeling Altruism and Spitefulness in Experiments. Review of Economic Dynamics 1(3): 593–622.

[57] Lucas, R. E. and Rapping, L. A. (1969). Price Expectations and the Phillips Curve. *The American Economic Review* 59(3): 342–350.

[58] Malmendier, U., and Schmidt, K. M. (2017) You Owe Me. American Economic Review. 107(2): 493–526.

[59] Mas, A., and Moretti, E. (2009). Peers at work. American Economic Review. 99(1): 112-145.

[60] Mauss, M. (1924). The Gift: The Form and Reason for Exchange in Archaic Societies. London: Routledge & Kegan Paul. (Translated by W. D. Halls. London: Routledge, 1990).

[61] Netzer, N., and Schmutzler, A. (2014). Explaining gift-exchange - The limits of good intentions. Journal of the European Economic Association.

[62] Owens, M. F., and Kagel, J. H. (2010). Minimum wage restrictions and employee effort in incomplete labor markets: An experimental investigation. Journal of Economic Behavior & Organization. 73(3), 317–326.

[63] Patel, A., and Smith, A. (2019) Guilt and participation. Journal of Economic Behavior & Organization 167: 279-295.

[64] Pearce, D. G. (1984). Rationalizable Strategic Behavior and the Problem of Perfection. Econometrica 52(4): 1029-1050

[65] Polonio, L., and Coricelli, G. (2018) Testing the level of consistency between choices and beliefs in games using eye-tracking. Games and Economic Behavior 113: 566-586

[66] Rabin M. (1993) Incorporating fairness into game theory and economics. The American economic review. 83(5): 1281–1302.

[67] Schotter, A., and Trevino, I. (2014). Belief Elicitation in the Laboratory. Annual Review of Economics 6: 103-128.