

Dhami, Sanjit; Wei, Mengxing; al-Nowaihi, Ali

Working Paper

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

CESifo Working Paper, No. 8992

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Dhami, Sanjit; Wei, Mengxing; al-Nowaihi, Ali (2021) : Classical and Belief-Based Gift Exchange Models: Theory and Evidence, CESifo Working Paper, No. 8992, Center for Economic Studies and Ifo Institute (CESifo), Munich

This Version is available at:

<https://hdl.handle.net/10419/235362>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

Sanjit Dhami, Mengxing Wei, Ali al-Nowaihi

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>

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

Abstract

We derive, compare, and test the predictions of three models of gift exchange: Classical (CGE); Augmented (AGE) based on unexpected wage surprises and first order beliefs; and Belief-based (BGE) that uses second order beliefs to formally model guilt-aversion. Motivated by Akerlof (1982), we also introduce signals of exogenous industry wage norms, θ_w , and effort norms, θ_e . We study the effects on the worker's optimal effort of exogenous variation in the wage, w , the signals θ_w , θ_e , and a signal of firm's expectations of effort from the worker, s . All three models successfully predict gift exchange (higher effort in response to higher w). The AGE and the CGE models fail to explain the data for the worker's effort responses to the signals θ_w , θ_e , and s . The BGE model successfully explains the data in all these respects. Gift exchange is underpinned by guilt-aversion and a formal modelling of second order conditional beliefs.

JEL-Codes: D010, D910.

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

30 March 2021

We are grateful to Martin Dufwenberg, Junaid Arshad, Yan Chen, and Stephanie W. Wang for comments and suggestions. 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, 6320202.

1 Introduction

The *gift exchange game* has played an important part in establishing the roles of *prosociality* and *reciprocity* in economics. The basic idea is that in a static game, if a worker is offered a binding wage by a firm in excess of the worker’s outside option, then the worker responds by putting in greater non-binding effort, 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.

In seminal experiments, Fehr et al. (1993, 1998) and Fehr et al. (1997) established that even controlling for the effect of confounding factors, *conditional reciprocity* to a gift was observed and has far reaching effects that include questions about the efficiency of competitive equilibria.¹ The gift exchange phenomenon survives large stakes experiments (Fehr et al., 2002); and 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). The powerful insights behind gift exchange also have a basis in an evolutionary account of human behavior and have led to important advances in contract theory, labour economics, and macroeconomic models among other areas (Bowles and Gintis, 2011; Dhami, 2019, Vol. 2).

Despite commendable progress in the understanding of conditional reciprocity, the underlying motivations behind gift exchange need further exploration. The purpose 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. The aim is to provide more rigorous microfoundations to underpin the motivations behind gift exchange.

Since we have three different models in the paper, a formal modelling of beliefs, and fairly stringent empirical tests of the theory, a somewhat structured approach is called for in the introduction. In Subsections 1.1–1.3 we briefly outline our 3 competing models; Subsection 1.4 explains our gift exchange game; Subsection 1.5 describes the transmission channels in all the models, provides intuition behind their predictions, and compares the predictions to the findings; Subsection 1.6 relates our paper to the literature; and Subsection 1.7 gives the section plans.

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 (e.g., Fehr et al., 1993, 1998 and Fehr et al., 1997) were motivated by the seminal work of Akerlof (1982). Workers who receive a wage, w , in excess of their

¹Netzer and Schmutzler (2014) show, purely in a theoretical model, that in situations where the firm’s wage increase cannot be interpreted as a kind intention on its part, then the worker may not respond with a positive gift exchange. Dufwenberg and Kirchsteiger (2019) show that these results are dependent on the efficiency definition that is used.

²Malmendier 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.”

reservation wage, reciprocate by putting in a higher effort, e ; the explanation given is in terms of *conditional reciprocity*.

Akerlof's (1982) original model also allows for a signal of exogenous wage norms (wages offered in similar firms in the industry), θ_w , to influence the worker's effort in a firm. But signals of *exogenous effort norms* (effort levels in similar firms in the industry), θ_e , do 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. We shall consider only exogenous wage and exogenous effort norms, and drop the prefix 'exogenous'.³

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 signals of wage norms, θ_w , and effort norms, θ_e . In classical gift exchange experiments, θ_w is 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 suggest a modification of the CGE model (Malmendier and Schmidt, 2017). If the worker receives an unexpected positive wage surprise, i.e., a wage in excess of his conditional 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. A similar idea has been implemented in several experimental papers on gift exchange (Gneezy and List, 2006; Kube et al., 2013). In this framework, effort reciprocity on the part of the worker is *gift-dependent*, but it is assumed to be *intentions-independent*; this is supported by the experimental results in Malmendier and Schmidt (2017). We call a model that includes a wage surprise term, an *AGE model* (Augmented Gift Exchange model).

It would appear unavoidable to use beliefs in gift exchange models. For instance, in the AGE model, workers use their first order beliefs to form wage expectations. However, in this paper, we classify belief-based models as those which require the essential use of *belief hierarchies*, in particular, second order beliefs (beliefs about the beliefs of other players). As we shall see below, this allows for a rigorous formulation of the emotions behind individual actions. This is not a matter of semantics. It also enables us to obtain new and testable implications that can discriminate between the alternative models.

1.3 Belief-based gift exchange (BGE)

Workers may experience a variety of emotions when they choose their optimal effort. In particular, we are interested in two kinds of emotions, *conditional reciprocity* and *guilt*; both may

³We do not directly consider endogenous wage/effort norms, which are norms established within a firm. It would appear that more foundational work is required on the nature of these endogenous norms before they can be formalized in theoretical models.

⁴Englmaier and Leider (2012) also propose a very 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.

include second order beliefs.⁵

1. Conditional reciprocity may trigger greater effort in response to a higher wage (gift exchange). In belief-based models, reciprocity could depend on workers' second order beliefs about the kindness intentions of the firm in setting a particular wage (intentions-dependent reciprocity).
2. 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 (i.e., the worker's second order beliefs about the first order beliefs of the firm).

The appropriate machinery to discuss guilt aversion and conditional reciprocity in a rigorous manner is provided by *psychological game theory*. In psychological game theory, unlike classical game theory, beliefs and belief hierarchies (i.e., beliefs about beliefs, such as second order beliefs) directly enter into the utility functions of players and these beliefs are endogenous (Geanakoplos et al., 1989; Rabin, 1993; Battigalli and Dufwenberg, 2009, 2020).

A large 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, 2020 and Dhami, 2019, Volume 4, sections 2.5, 3.6). Considerations of conditional reciprocity in psychological game theory were introduced by Rabin (1993), extended to extensive form games by Dufwenberg and Kirchsteiger (2004), and to an even more general framework (that includes reciprocity) by Battigalli and Dufwenberg (2009). Considerations of guilt were introduced by Battigalli and Dufwenberg (2007) and we shall be mainly interested in what they term as *simple guilt*.⁶

Guilt also lies behind many other explanations of human behavior including moral norms and the norm of reciprocity (Bicchieri, 2006; Elster, 2011) and the applications have increased considerably in recent years (Battigalli and Dufwenberg, 2020 and Dhami, 2019, Volume 4). We term models that rest on belief hierarchies, as *belief-based gift exchange* (BGE) models. However, we will not need beliefs of an order higher than second order beliefs in this paper.⁷

1.4 Our gift exchange experiments

Prior to choosing the wage, w , the firm receives signals of the typical *wage* (θ_w) and *effort* (θ_e) in similar firms.⁸ Second, we also elicit a signal, s , of the first order belief of the firm about the

⁵With the exception of Dhami et al. (2019a), 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.

⁶Another concept of guilt aversion, *guilt from blame*, requires the formation of third and fourth order beliefs (Battigalli and Dufwenberg, 2007). For applications of this concept and the empirical evidence, see Khalmetski et al. (2015) and Dhami et al. (2019a). 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.

⁷Dhami et al. (2019a) show, for instance, that fourth order beliefs play a role in determining public goods contributions in public goods games. We leave an exploration of higher order beliefs for the gift exchange game for future research.

⁸By highlighting the word 'typical' in our experiments, we hope to bring about a *shared understanding* of the wage and effort norms in similar firms; shared understanding is essential for the establishment of successful norms (Fehr and Schurtenberger, 2018).

effort level it expects from the worker. Firms then decide on the wage w to be offered to the worker. Workers are split into two types; the signal s is revealed to the *informed workers* but not to the *uninformed workers*. Both types of workers are informed of the wage, w , and the signals θ_w, θ_e . Thus, the respective information sets of the informed and uninformed workers are $\{w, \theta_w, \theta_e, s\}$ and $\{w, \theta_w, \theta_e\}$ at the instant when they choose their effort e .

We seek 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 from various disciplines to test the comparative statics with respect to w, θ_w, θ_e . Subjects were randomly assigned to the roles of firm, informed worker, and uninformed worker. In total, and using the strategy method, across all sub-treatments we collected 12,144 data points for decisions made by firms and workers. A set of further, online, experiments were conducted in June-July 2020 in China with a similar subject pool in order to test the comparative statics with respect to s (see Section 7 for the details).

1.5 Transmission mechanisms, predictions, and findings

All three models, CGE, AGE and BGE, predict an increase in the worker's optimal effort, e^* , in response to an increase in w (gift exchange), in line with Akerlof (1982) and our empirical evidence. Our empirical evidence also shows an increase in effort in response to increases in θ_w, θ_e and, for the informed worker, s . The latter is predicted only by the BGE model. By contrast, the other two models, CGE and AGE, predict the wrong comparative static results with respect to θ_w, θ_e, s .

In the CGE model, if workers observe that the firm they are matched with, pays a wage w greater than the signal of wage norm in similar firms, θ_w , they put in effort e greater than the signal of effort norm in similar firms, θ_e . The higher is θ_w , the lower is the size of the perceived gift by the matched firm, so optimal effort is reduced (contrary to our evidence). However, the effort norm signal, θ_e , has no effect on optimal effort in the CGE model (as in Akerlof, 1982). This is because, although θ_e enters the worker's utility function, it does not enter the marginal utility. The signal s is predicted to have no effect on the optimal effort choice because there is no guilt aversion channel. The signal s also plays no role in Akerlof (1982).

In the AGE model, the main transmission channel is unexpected wage changes. Expected wage, in turn, is conditioned on the relevant information sets of the workers, namely, the signals θ_w, θ_e, s . Using a plausible first order stochastic dominance assumption, an increase in any of these signals increases the wage expectation of the worker. This reduces the gift, as perceived by the worker, who therefore reduces optimal effort.

In the BGE model, guilt-averse workers feel *guilty* from letting down the firm's effort expectations, inducing them to exert higher effort. Workers use their signals (θ_w, θ_e, s) , to form more precise inferences about the firm's first order beliefs (*worker's second order beliefs conditional on the signals*). A worker who receives a higher signal s reasonably believe that the firm expects higher effort from him.⁹ An increase in the other two signals, θ_w and θ_e , also increase

⁹Indeed, in the employment relationship, firms often convey their effort expectations to workers in a variety

the likelihood that the firm expects a higher effort from the worker.¹⁰ Thus, an increase in any of θ_w, θ_e, s increases the worker’s belief that the firm expects a high effort level. Guilt averse workers, must then increase their optimal effort so as not to fall below the firm’s expectation. Our empirical data strongly confirm all four comparative static predictions of the BGE model with respect to w, θ_w, θ_e, s . By contrast, the comparative static predictions of the CGE and AGE models with respect to θ_w, θ_e, s are rejected.

To be sure, existing empirical evidence supports gift exchange, even when the signals θ_w, θ_e, s are absent (conditional reciprocity). This was sufficient to make the important point that exclusive reliance on self-regarding preferences needs to be modified. The gift exchange game is critical to an understanding of the employment relation, and despite many fruitful extensions (see below), further avenues based on Akerlof’s (1982) treatment of gift exchange need to be explored. A range of emotions (e.g., guilt aversion and intentions based reciprocity), and factors (e.g., industry wage/effort norms), mediate the gift exchange relation. In this sense, we contribute towards a more fuller understanding of gift exchange.

1.6 Relation to the literature

The gift exchange relation can also be 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 .¹¹

The early literature on testing for guilt-aversion used *direct belief elicitation* and concluded that guilt aversion was important.¹² For instance, in a two player game, e.g., trust game or the ultimatum game, in order to elicit *player 1’s second order beliefs*, player 1 was directly asked to guess what player 2 believed about player 1’s beliefs. Such *direct belief elicitation* is heavily prone to the *false consensus effect*, namely, that players assign their own beliefs to others, which potentially invalidates tests based on such elicitation methods. For this reason, we follow an amended version of the *induced beliefs design* of Ellingsen et al. (2010) that eliminates the false consensus effect. Great care is taken to ensure that the firm does not misstate its of ways (mission/policy statements; firm culture; periodic manager-worker or union-firm meetings).

¹⁰If the effort norm, θ_e , in the industry is high, then the worker may reasonably expect the firm to expect high effort from its workers. If the wage norm, θ_w , is high, then given that gift exchange is understood by the players, workers should reasonably expect that this creates an expectation of a higher effort. We formalize these insights by assuming that the distribution of second order beliefs of the worker has the first order stochastic dominance property with respect to the variables θ_w, θ_e, s .

¹¹This 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, 2020 for the formal definitions) may also be able to provide an alternative explanation for the existence of punishments within the domain of our BGE model.

¹²Since our study does not use the direct belief elicitation method, we omit the large number of references. The necessary references can be found in footnote 75, p. 48 in Battigalli and Dufwenberg (2020). See also the literature survey in Cartwright (2019) and the references in Dhimi et al. (2019a).

beliefs strategically; that there is no subject deception; that this is mutual knowledge to the worker; and our use of the strategy method for eliciting optimal effort responses to s should serve to eliminate any residual concerns with this method (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.¹³

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 Dhimi (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, the relevant emotions such as reciprocity and guilt, and the response to the signals θ_w, θ_e, s , cannot be studied using a formal and rigorous belief-based framework.

1.7 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 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 qualitative results can be generalized, but we suppress generality to keep model as close as possible to our experimental design.

¹³For instance, guilt aversion explains greater dictator giving when the receiver expects more (Khalmetzki et al., 2015; Hauge, 2016); has a critical influence on public good contributions in public goods games (Dhimi et al., 2019a); influences embezzlement 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 (Dhimi et al., 2019b). For surveys, see Battigalli and Dufwenberg (2020), Cartwright (2019), and Dhimi (2019, Vol. 4). For general issues in belief elicitation that go beyond psychological game theory, see Schotter and Trevino (2014).

2.1 Preferences and technology

Consider a standard gift exchange game between one firm (F) and one worker (W); both are risk neutral. 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 function¹⁴. Thus, we adopt the profit function

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

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

The *material utility* of the worker is

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

where $20e^2$ is the disutility of work to the worker. Effort is not contractually binding, hence, it cannot be part of the formal employment contract.

2.2 Beliefs and sequence of moves

The generic gift exchange game has two stages. In Stage 1 the firm chooses a contractually binding wage, w . Then, in Stage 2, the worker chooses non-contractible effort, e . 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 some initial beliefs, that we describe below.

First order beliefs: These are beliefs about the choices that one expects other players to make. In particular, b_F^1 denotes the first order belief of the firm about the effort, $e \in [0, 1]$, to be exerted by the worker. Analogously, b_W^1 is the first order belief of the worker about the wage, $w \in [0, 100]$, to be paid by the firm. The cumulative probability distributions of b_F^1 and b_W^1 are, respectively, $P_F^1(e)$ and $P_W^1(w)$.¹⁵

Second order beliefs: These are beliefs of the players about the first order beliefs of other players and allow for a rigorous treatment of *guilt*. In particular, b_W^2 is the second order belief of the worker about the firm's first order belief, b_F^1 , about effort $e \in [0, 1]$. The unconditional cumulative probability distribution of b_W^2 is $P_W^2(e)$.¹⁶

¹⁴For 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.

¹⁵A 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.

¹⁶We do not need to specify the second order belief of the firm, b_F^2 , as they play no role in our analysis.

2.2.1 Stage 1 (Firms: Belief updating and wage setting)

In Stage 1, the firm chooses the contractible wage, w , but after updating its beliefs, as we now explain.

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. After observing these signals, but before choosing the wage, w , the firm updates its initial belief b_F^1 about the likely effort that the worker will exert. The updated cumulative distribution function of b_F^1 is $P_F^1(e|\theta_w, \theta_e)$.

Example 1 : *Suppose that the firm receives a high signal, θ_e , of effort in similar firms. Then, the firm may assign a high probability that the worker it has employed, and who also observes θ_e , will work hard. This is encapsulated in the firm's first order conditional belief distribution $P_F^1(e|\theta_w, \theta_e)$.*

Once the firm has observed θ_e and θ_w , the experimenter elicits from the firm a signal, $s \in [0, 1]$, of the firm's first order belief, b_F^1 , about the effort level, e , expected from the worker. The signal s is a point estimate made by the firm, based on the underlying distribution of its beliefs, $P_F^1(e|\theta_w, \theta_e)$. For example, s could be mean, median, or mode of $P_F^1(e|\theta_w, \theta_e)$.

Finally, the firm chooses the contractible wage level $w \in [0, 100]$ for the worker. This concludes Stage 1.

2.2.2 Stage 2 (Workers: Belief updating and choice of effort)

In Stage 2, the worker chooses the effort level, e , but after observing the contractible wage, w , and the signals of industry effort and wage norms, θ_e and θ_w , which allow the worker to update his/her beliefs. 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 (Treatment 1). Half the workers are in each treatment.

Definition 1 : *Define Γ_j to be the information set containing all the signals, but excluding the wage, w , of the worker of type $j = I, N$, where I refers to a worker who has been informed of the signal $s \in [0, 1]$ of the effort level, e , the firm expects from the worker. N refers to a worker who has not been informed. In particular,*

$$\Gamma_I = \{\theta_w, \theta_e, s\} \text{ and } \Gamma_N = \{\theta_w, \theta_e\}. \quad (2.3)$$

Let $\bar{\gamma} = 100$, if $\gamma = \theta_w$, and let $\bar{\gamma} = 1$, if $\gamma \in \{\theta_e, s\}$.¹⁷

Using the signals, θ_w, θ_e, s , workers update their initial or first order beliefs b_W^1 (about the wage, w , paid by the firm) and the second order beliefs b_W^2 (about the firm's first order belief, b_F^1 , about effort to be exerted by the worker). Then the cumulative distributions of the updated beliefs are given by $P_W^1(w|\Gamma_j)$ and $P_W^2(e|\Gamma_j)$, $j = I, N$.

¹⁷In order to define first order stochastic dominance (see, e.g., Assumption A3 below) we need to define the domains of the relevant variables in Γ_j . Since wage $w \in [0, 100]$, thus, the signal of wage in similar firms $\theta_w \in [0, 100]$, hence the upper limit $\bar{\gamma} = 100$. On the other hand, 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 $\bar{\gamma} = 1$

Example 2 : Suppose that the signal θ_e , of effort levels in similar firms, is high. In the light of this signal, the worker may assign a high probability that the firm expects a high effort from its employed workers (encapsulated in the worker's second order cumulative belief distribution $P_W^2(e|\Gamma_j)$, $j = I, N$).

In all our treatments and sub-treatments, to be described later, the effort level, e , is chosen by the worker, conditional on the relevant signals θ_w, θ_e, s as well as the wage w announced by the firm.

2.3 Assumptions on beliefs

We are primarily interested in the cumulative belief distribution functions of workers.

Assumption A1 (Continuity) Beliefs are continuously distributed. It follows that $P_W^1(w)$ and $P_W^2(e)$ are integrable.

Assumption A2 (Differentiability) The first and second order cumulative belief distribution functions $P_W^1(w|\Gamma_j)$ and $P_W^2(e|\Gamma_j)$, $j = I, N$, are differentiable with respect to $\gamma \in \Gamma_j$.

We shall make two further assumptions on the belief distributions, A3 in Section 5 (the AGE model), and A4 in Section 6 (the BGE model).

2.4 A note on best response to beliefs

The main solution concept in psychological game theory is a *sequential equilibrium*. Essentially (and as in classical game theory) actions are optimal given beliefs, and beliefs of all orders are correct given actions (the assumption of *rational expectations*). There are two problems with this approach; the second problem (non-conformity with the evidence) is more serious.

1. Multiple equilibria often arise in psychological game theory models that impose sequential equilibrium (Rabin, 1993; Fehr and Schmidt, 2006). Malmendier and Schmidt (2017) show that, within the gift exchange game, this allows for the presence as well as the absence of gift exchange; which makes it difficult to test the theory.
2. Contrary to the rational expectations assumption that underpins the sequential equilibrium, there is much evidence for disequilibrium beliefs, particularly in games played only a few times. Many behavioral solution concepts are also based on the idea of disequilibrium beliefs, such as the level-k model, the cognitive hierarchy model, evidential equilibrium, and analogy based equilibria (Dhami, 2019, Vol. 4). 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.

For these reasons, several recent papers using psychological game theory, like our paper, 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*. The predictions of these models, particularly that use psychological game theory, match well with the evidence (Khalmetzki et al., 2015; Dhama et al., 2019a; Dhama et al., 2019b). The problem of multiple equilibria is avoided in this literature.

3 Classical gift exchange (CGE)

In the classical gift exchange game, for a given wage, w , the worker chooses his optimal effort, $e = e^*$, to maximize the following utility function.

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

The first two terms on the RHS of (3.1) give the worker’s material payoff. The third term, $\beta(w - \theta_w)(e - \theta_e)$, 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 workers could receive if they left their current jobs*. We believe that (3.1) is closer to Akerlof’s original formulation of gift exchange.¹⁸

The third term in (3.1) is the essence of gift exchange in this *intentions-independent* framework. Note that the signal, s , plays no role in this formulation of gift exchange 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 $\frac{\partial e^*(w, \Gamma_j)}{\partial w} > 0$, $\frac{\partial e^*(w, \Gamma_j)}{\partial \theta_w} < 0$.
- (c) If $w \geq \frac{40}{\beta} + \theta_w$, then $e^*(w, \Gamma_j) = 1$.
- (d) In all cases, $\frac{\partial e^*(w, \Gamma_j)}{\partial \theta_e} = \frac{\partial e^*(w, \Gamma_j)}{\partial s} = 0$.

From Proposition 1b optimal effort is increasing in the wage rate, w (gift exchange). However, 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. Note that neither θ_e nor s have an effect on the optimal effort level, $e^*(w, \Gamma_j)$.

Thus, 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, it also predicts $\frac{\partial e^*(w, \Gamma_j)}{\partial \theta_w} < 0$ and $\frac{\partial e^*(w, \Gamma_j)}{\partial \theta_e} = \frac{\partial e^*(w, \Gamma_j)}{\partial s} = 0$, which are all rejected by our data.

¹⁸In 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 $\beta w e$.

4 Augmented gift exchange (AGE)

The modern literature on gift exchange retains the central gift exchange feature of the CGE model but brings in 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.”

Remark 1 : Central to the Malmendier-Schmidt formulation is the idea that “... the more the favorable act exceeds expectation, the stronger the positive response.” Thus, economic actors react to the surprise or the disappointment that is generated by the difference between their prior beliefs, as captured in their first order cumulative belief distribution (in our case, for the workers, $P_W^1(w|\Gamma_j)$, $j = I, N$) and the actual actions of the other players (in our case, the wage w chosen by the firm). Thus, it becomes important to specify how $P_W^1(w|\Gamma_j)$ is influenced by parameters in Γ_j . This is done below in Assumption A3, which is eminently reasonable, as the discussion following it shows.¹⁹

In our *augmented gift exchange* (AGE) model we propose that the utility of the worker in the presence of reciprocity, that arises from (positive or negative) wage surprises, is given by

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

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

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

The first two terms on the RHS of (4.1) give the material payoff of the worker. The third term, $\frac{\sigma}{100} [w - E(w|\Gamma_j)] \pi(e, w)$, is the reciprocity payoff to the worker based on positive ($w > E(w|\Gamma_j)$) or negative ($w < E(w|\Gamma_j)$) wage surprises.²⁰ Thus, the firm’s profits, $\pi(e, w)$, can be internalized positively or negatively in the worker’s utility function. 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.

¹⁹This is not to imply that the role of expectations is not recognized in Akerlof (1982). However, the word ‘expectations’ appears explicitly only once in Akerlof (1982) and it is not the central focus. Neither are expectations the focus in the CGE model (although they are implicit in the term θ_w, θ_e in (3.1)). By contrast, the formation of expectations and belief hierarchies is central in the BGE model in Section 5.

²⁰This 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.

Integrating (4.2) by parts, gives

$$E(w|\Gamma_j) = 100 - \int_{w=0}^{100} P_W^1(w|\Gamma_j) dw. \quad (4.3)$$

Substitute from (4.3) into (4.1) and use (2.1) to get

$$V(e; w, \Gamma_j) = w - 20e^2 + \frac{\sigma}{100} (100e - w + \kappa) \left[w - 100 + \int_{w=0}^{100} P_W^1(w|\Gamma_j) dw \right], \quad (4.4)$$

where $\sigma > 0$ and $j = I, N$. The factor $\frac{1}{100}$ in (4.1) and (4.4) is introduced for computational convenience, without affecting any results.

Assumption A3 (Strict first order stochastic dominance for first order beliefs of a worker) For any $\gamma \in \Gamma_j$, $j = I, N$, a higher value of γ induces strict first order stochastic dominance in $P_W^1(w|\Gamma_j)$. If $\gamma_1 < \gamma_2$ are two distinct values of $\gamma \in \Gamma_j$, then we have $P_W^1(w|\gamma_2, \cdot) < P_W^1(w|\gamma_1, \cdot)$.

Since beliefs are differentiable (Assumption A2), and using Definition 1, Assumption A3 implies that

$$\frac{\partial P_W^1(w|\Gamma_j)}{\partial \gamma} < 0. \text{ for all } \gamma \in \Gamma_j, 0 < \gamma < \bar{\gamma}, j = I, N. \quad (4.5)$$

Discussion of Assumption A3: Assumption A3 requires that higher values of $\gamma \in \Gamma_j$ increase the probability, in the worker's mind, prior to observing the wage in Stage 2, that the firm will offer a high wage. To build intuition around Assumption A3, it is useful to assume that, since reciprocity is ubiquitous, players assign some likelihood that other players might be reciprocal. (i) Consider θ_w . In firms with a high wage norm θ_w it is reasonable for the worker to initially expect a high wage from the firm. (ii) Consider $s \in \Gamma_I$. If the worker believes that the firm is reciprocal, then a higher signal s from the firm about its expectations of worker's effort makes it less likely that the firm will offer a low wage. (iii) Consider θ_e . A high effort norm in similar firms makes it reasonable for the worker to expect that the paired firm also expects a high effort. Since workers expect the firm to be reciprocal, they expect that the firm will reciprocally offer a high wage.

Proposition 2: The worker's utility (4.4) has a unique maximum, $e^*(w, \Gamma_j)$. Under Assumption A3 it has the following properties:

- (a) $\frac{\partial e^*(w, \Gamma_N)}{\partial s} = 0$.
- (b) If $w \leq E(w|\Gamma_j)$, then $e^*(w, \Gamma_j) = 0$.
- (c) If $E(w|\Gamma_j) < w < \frac{40}{\sigma} + E(w|\Gamma_j)$, 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) If $w \geq \frac{40}{\sigma} + E(w|\Gamma_j)$, then $e^*(w, \Gamma_j) = 1$.

Discussion of Proposition 2: In the AGE model (as in the CGE model) $\frac{\partial e^*(w, \Gamma_j)}{\partial w} > 0$, so both models capture the essence of gift exchange, which is in line with our evidence. However,

the AGE model also predicts that $\frac{\partial e^*(w, \Gamma_j)}{\partial \theta_w} < 0$ and $\frac{\partial e^*(w, \Gamma_j)}{\partial \theta_e} < 0$. For the informed worker, it predicts $\frac{\partial e^*(w, \Gamma_j)}{\partial s} < 0$. The intuition behind the comparative static results with respect to θ_w , θ_e , and s is that an increase in any of these signals increases the wage expectation of the worker (see Assumption A3). This reduces the gift, as perceived by the worker, who optimally reduces effort. All three predictions with respect to the signals are contrary to our evidence.

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

As noted in the introduction, there is now extensive and compelling evidence that *beliefs* and *beliefs about beliefs* directly enter into the utility function of individuals. This is an essential feature in models of psychological game theory, and it enables the modelling of a range of emotions. 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

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

where

$$\lambda_R > 0, \lambda_G > \lambda_S \geq 0, \quad (5.2)$$

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) will turn out to be 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:

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

From (5.3), the worker faces disutility on account of guilt if his actual effort, e , falls short of what he believes is the firm's expectation of his effort level. Since the worker does not directly observe the firm's expectations (or the first order belief of the firm, b_F^1), he uses his own second order beliefs, b_W^2 , which are his beliefs about b_F^1 . The cumulative probability function of b_W^2 , conditional on the available signals, is $P_W^2(e|\Gamma_j)$, $j = I, N$.

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; Dhimi et al. 2019a):

$$S(e, \Gamma_j) = \int_{x=0}^{x=e} (e - x) dP_W^2(x|\Gamma_j), \quad j = I, N. \quad (5.4)$$

Khalmetski et al. (2015) and Dhimi et al. (2019a) explicitly measure the relative sizes of λ_G , λ_S 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). Therefore, we shall refer to the combination $-\lambda_G G(e, \Gamma_j) + \lambda_S S(e, \Gamma_j)$ in (5.1) as the *guilt-aversion channel*.

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

$$G(e, \Gamma_j) = 1 - e - \int_{x=e}^{x=1} P_W^2(x|\Gamma_j) dx, \quad (5.5)$$

$$S(e, \Gamma_j) = \int_{x=0}^{x=e} P_W^2(x|\Gamma_j) dx, \quad j = I, N. \quad (5.6)$$

Reciprocity (R): We apply the framework of Dufwenberg and Kirchsteiger (2004) for sequential games to specify the conditional reciprocity term $R(e, w, \Gamma_j)$ in (5.1).²¹ In this framework,

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

where $k_{WF}(e, w)$ is the *kindness of the worker to the firm, as perceived by the worker* and $\widehat{k}_{FW}(w, \Gamma_j)$ 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 ($k_{WF} < 0$). The essential idea behind kindness functions is to compare the *expected payoff of a player* to a player's *equitable payoff* that captures some notion of a fair payoff. The equitable payoff, in turn, is a weighted average of the maximum and the minimum payoffs that accrue to a player, from the actions taken by the other player. This motivates Definition 2, below.

Definition 2 (Dufwenberg and Kirchsteiger, 2004):

(a) The equitable payoff of the firm, $\pi^E(w)$, as perceived by the worker, is

$$\pi^E(w) = \mu \max\{\pi(e, w), e \in [0, 1]\} + (1 - \mu) \min\{\pi(e, w), e \in [0, 1]\},$$

where $\pi(e, w)$ is given by (2.1) and $\mu \in [0, 1]$.

(b) The kindness of the worker to the firm, $k_{WF}(e, w)$, as perceived by the worker, is

$$k_{WF}(e, w) = \pi(e, w) - \pi^E(w).$$

(c) Let $Eu(w, \Gamma_j) = \int_{e=0}^1 u(e, w) dP_W^2(e|\Gamma_j)$, where $u(e, w)$ is given by (2.2), and $P_W^2(e|\Gamma_j)$ is the cumulative probability distribution of b_W^2 . The latter is the second order belief of the worker about the first order belief, b_F^1 , of the firm about the worker's effort level, e .

(d) The equitable payoff to the worker, $u^E(w, \Gamma_j)$, as perceived by the worker, is

$$u^E(w, \Gamma_j) = \nu \max\{Eu(w, \Gamma_j), w \in [0, 100]\} + (1 - \nu) \min\{Eu(w, \Gamma_j), w \in [0, 100]\},$$

where $u(e, w)$ is given by (2.2) and $\nu \in [0, 1]$.

(e) The kindness of the firm to the worker, $\widehat{k}_{FW}(w, \Gamma_j)$, as perceived by the worker, is

$$\widehat{k}_{FW}(w, \Gamma_j) = Eu(w, \Gamma_j) - u^E(w, \Gamma_j).$$

(f) The worker's conditional reciprocity towards the firm, $R(e, w, \Gamma_j)$, is

$$R(e, w, \Gamma_j) = k_{WF}(e, w) \widehat{k}_{FW}(w, \Gamma_j).$$

Lemma 1 : The worker's conditional reciprocity towards the firm, $R(e, w, \Gamma_j)$, is given by

$$R(e, w, \Gamma_j) = 100(w - 100\nu)(e - \mu), \quad (5.8)$$

²¹For an analysis of alternative methods of modelling kindness functions, such as those based on the seminal model by Rabin (1993), see Dufwenberg and Kirchsteiger (2019) and Dhami (2019; Vol. 4, Section 2.5).

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

The parameter μ captures 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 captures 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 $(w - 100\nu)(e - \mu)$. If 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 he 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:

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

Remark 2 (*Beliefs in the competing models*): In the CGE model (3.1), we had neither first order nor second order beliefs. In the AGE model (4.4), first order beliefs entered into the objective function through wage expectations 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 Lemma 1 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: (First order stochastic dominance for second order beliefs of a worker) A higher value of the signal $\gamma \in \Gamma_j$, $j = I, N$, induces strict first order stochastic dominance in the conditional distribution of beliefs $P_W^2(e|\Gamma_j)$. If $\gamma_1 < \gamma_2$ are two distinct values of $\gamma \in \Gamma_j$, then for all $\theta_w \in (0, 100)$ and $e, s, \theta_e \in (0, 1)$ we have $P_W^2(e|\gamma_2, \cdot) < P_W^2(e|\gamma_1, \cdot)$.

Since beliefs are differentiable (Assumption A2), Assumption A4 implies that, for all $\theta_w \in (0, 100)$ and $e, \theta_e, s \in (0, 1)$:

$$\frac{\partial P_W^2(e|\Gamma_j)}{\partial \gamma} < 0, \quad \gamma \in \Gamma_j, \quad j = I, N. \quad (5.10)$$

Discussion of Assumption A4: $P_W^2(e|\Gamma_j)$ is the cumulative probability distribution of second order beliefs of the worker, which gives the worker's best guess of how much effort the firm

expects from him. 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. In firms with a high wage norm θ_w , it is reasonable for the firm to expect that workers will put in more effort. Thus, the worker would assign higher probability, in terms of his second order beliefs, that the firm is more likely to expect higher effort from him. A high effort norm in similar firms makes it reasonable for the worker to expect that the matched firm also expects a high effort from him. A higher value of s offers direct evidence to the informed worker that the firm expects higher effort from him.

Proposition 3: 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 $100\nu - \frac{\lambda_G}{\lambda_R} < w < 100\nu + \frac{40}{\lambda_R}$. 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) $\frac{\partial e^*(w, \Gamma_j)}{\partial s} > 0$, for $s \in (0, 1)$.
- (d) Suppose $w \geq 100\nu + \frac{40}{\lambda_R}$. Then $e^*(w, \Gamma_j) = 1$.

At an interior solution, 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 3 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). Hence, guilt-averse/surprise-seeking workers try to put in higher effort in order to reduce the possibility of ex-post guilt and increase the possibility of ex-post elation.

Remark 3 : 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 $\frac{\partial e^*(w, \Gamma_j)}{\partial \theta_w} = \frac{\partial e^*(w, \Gamma_j)}{\partial \theta_e} = 0$.

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 1: A summary of the comparative static results under the different models when the optimal effort level is an interior point

Table 1 compares the predictions of the AGE and the CGE models (Propositions 1, 2) with the BGE model (Proposition 3) 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 discriminate between the models.

7 Experimental Design

In subsection 7.1, below, we describe lab experiments we performed 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 1, 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 1. 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 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. 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. The wage-effort choices in matched pairs of firm-worker subjects generated actual payoffs. One of the choices was picked at random and paid.

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 given the respective material profit and utility functions 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 0.05.²² 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

²²Subjects in the experiment could choose any effort level between 0 and 1.

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 in the role of informed workers, and 96 subjects in the role of uninformed workers. There were three different treatments for each worker, where in each treatment a worker made 21 choices. Following the strategy method, we 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. 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.²³ Then they 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 a 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 then, in Stage 2, chose a non-contractible effort, e . Stage 2 had two treatments, and each treatment had three sub-treatments. We describe these below.

Treatment 1 (The uninformed worker): In order to test for the 3 comparative static effects with respect to w, θ_w, θ_e in columns 2, 3, 4, respectively, in Table 1, we conducted three sub-treatments A, B, C. 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.²⁴ 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.

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

²³Our 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).

²⁴We 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 report below in Section 8.5, when we varied the order for one of the comparative static effects (with respect to θ_w) there was no change in the results.

strategy method to determine the effort response $e \in [0, 1]$ of each worker to 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, $\frac{\partial e^*}{\partial w}$. If this sub-treatment is played for real (after being randomly chosen) then the actual wage chosen by the firm and the corresponding effort chosen by the worker in the strategy method are paired to determine the payoffs.

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, $\frac{\partial e^*}{\partial \theta_w}$. If this sub-treatment is played for real (after being randomly chosen) then one of the randomly chosen 21 equally spaced hypothetical wage norm signals is used to determine the payoffs.

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 *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 this sub-treatment is played for real (after being randomly chosen) then one of the randomly chosen 21 equally spaced hypothetical effort norm signals is used to determine the payoffs.

Treatment 2 (The informed worker): Following the induced beliefs design (Ellingsen et al., 2010; Khalmetski et al., 2015; Dhami et al., 2019a), 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 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, Treatment 2 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 sub-treatment, sub-treatment 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 the lab experiments ($w = 53$) and identical values $\theta_w = 40$ and $\theta_e = 0.4$ for the 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 give clear guidance on 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.²⁵

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 2: 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 2 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 1. We first report the correlation between effort and the relevant variable, for each individual worker, and then we report the histogram of regression coefficients across all worker, 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)

In Table 3 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. As explained

²⁵We ran the online experiments using Wenjuanxing, which is a widely used Chinese platform providing functions that are equivalent to Amazon Mechanical Turk.

in Section 7, we obtain the optimal effort data for each worker by varying the wage rate, w , between 0 and 100 in increments of 5 for 21 different values of w . For the 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.

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

Table 3: Number and percent of the Spearman correlation coefficients between wage and effort which are significant at the 1 percent level.

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.

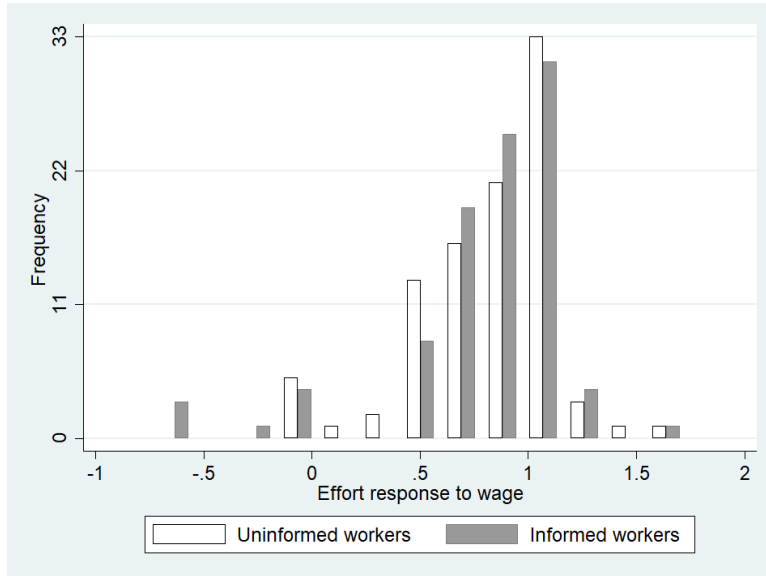


Figure 1: Histogram of regression coefficients, significant at 1%, when optimal effort is regressed on wage, w , for informed and uninformed workers.

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

Table 4 shows the Spearman correlation coefficients between effort, e , and the signal of effort norm, θ_e , that are significant at the 1% level. We obtain the optimal effort data for each worker by varying θ_e , between 0 and 1 in increments of 0.05 for 21 different values of θ_e (see Section 7). The great majority of the subjects exhibit $\rho(e, \theta_e) > 0$. From Table 1, the prediction of the CGE model is that $\frac{\partial e^*}{\partial \theta_e} = 0$ and this is also the prediction of the AGE model; this prediction

holds for 5.2% of the subjects, but it is rejected for the rest. 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 4: Number and percent of the Spearman correlation coefficients between effort and effort norm which are significant at the 1 percent level.

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.

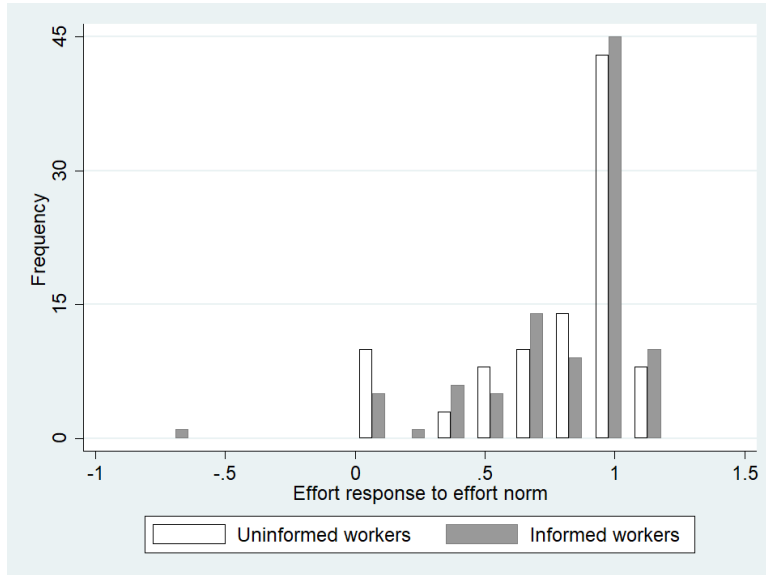


Figure 2: Histogram of regression coefficients significant at 1% when effort is regressed on the signal of effort norm, θ_e , for informed and uninformed workers.

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

Table 5 shows the Spearman correlation coefficients between effort, e , and the signal of wage norm, θ_w , that are significant at the 1% level. We obtain the optimal effort data for each worker by varying θ_w , between 0 and 100 in increments of 5 for 21 different values of θ_w (see Section 7). 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 1). By contrast, the prediction of the CGE and AGE models is that $\frac{\partial e^*}{\partial \theta_w} < 0$, which holds for 17.7% of the subjects.

Figure 3 shows the histogram of the regression coefficients that are significant at 1%, when we regress effort on the wage norm 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

	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 5: Number and percent of the Spearman correlation coefficients between effort and wage norm which are significant at the 1 percent level.

consistent with the CGE and AGE models. However, the vast majority of the informed and uninformed workers have significantly positive coefficients, which is consistent with the BGE model.²⁶

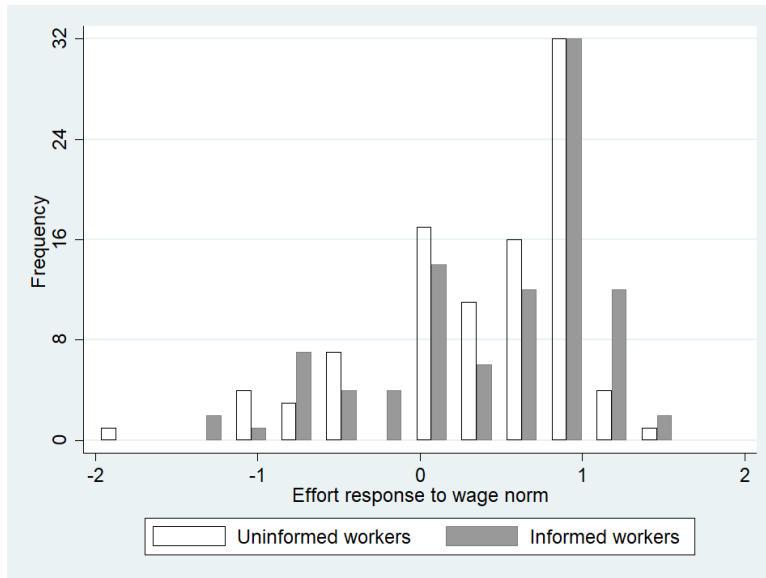


Figure 3: Histogram of regression coefficients significant at 1% when effort is regressed on the signal of wage norm, θ_w , for informed and uninformed workers.

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, 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 $\rho(e, \theta_w) > 0$, which is very close to the figure of 68.8% in Table 5. This provides some 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

²⁶In Tables 3, 4, one informed subject and one uninformed subject chose zero effort level at all levels of the independent variables; their data was excluded. Similarly in Table 5, one informed and two uninformed subjects chose zero effort level for all values of the independent variable, so their data was excluded as well.

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 from 6 subjects, because they did not answer the control questions correctly, hence, we report the results for the remaining 52 subjects.²⁷

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 highly positive and significant at the 1% significance level. 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 AGE and the CGE models (see column 5 in Table 1).

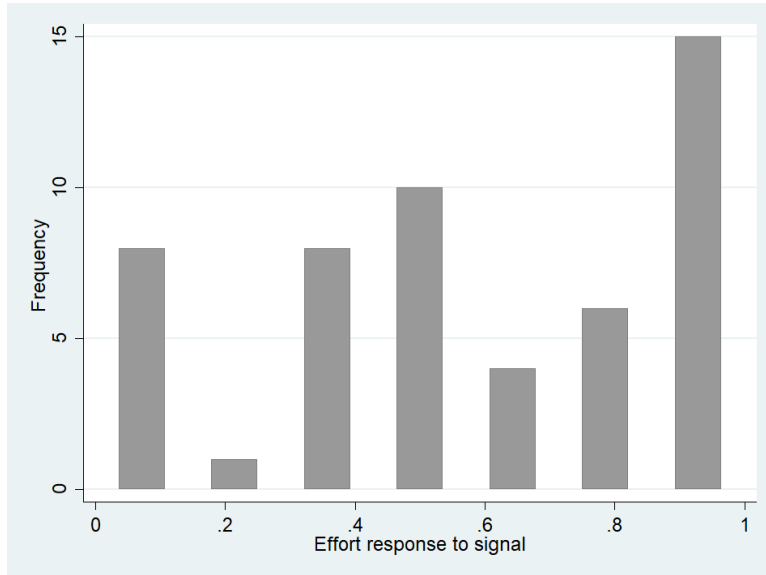


Figure 4: Histogram of regression coefficients, significant at 1%, when optimal effort of informed workers is regressed on the signal, s .

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 significantly positive regression coefficient (at the 1% significance level), 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,²⁸ the average size of the regression coefficients is 0.6.

²⁷The 52 subjects spent on average of 9 minutes completing the experiment, and they were paid an average of around 14 Yuan.

²⁸Among the 10 cases with non-monotonic effort responses to the signals, 8 of them exerted monotonically decreasing efforts at some value of the signals.

In sum, the statistical results strongly support the prediction of the BGE model but not the other two models.

9 Determinants of effort

We now consider the determinants of effort choice in several Tobit models (censored on both sides) in Table 6, 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 6: Determinants of effort.

	Dependent Variable: Effort					
	Informed Workers				Uninformed Workers	
Tobit Model	1	2	3	4	5	6
Signal	0.399*** [0.077]	0.149** [0.074]	0.367*** [0.094]	0.151* [0.082]		
Wage		0.006*** [0.001]		0.005*** [0.001]	0.011*** [0.001]	0.011*** [0.001]
Male			0.075** [0.032]	0.051 [0.034]		-0.020 [0.052]
Education			0.030*** [0.012]	0.019 [0.012]		0.018 [0.019]
Experience			-0.019 [0.069]	-0.007 [0.077]		0.028 [0.048]
Constant	0.232*** [0.058]	0.061 [0.054]	0.162*** [0.055]	0.035 [0.054]	-0.121** [0.047]	-0.154 [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 percent, 5 percent, and 10 percent, 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 $\{1, 2, \dots, 7\}$ and higher values denote higher educational attainment, e.g., Education = 1 for first year undergraduate students and Education = 6 for second year master students²⁹.

In the experiments, the uninformed group did not receive the signal s of the firm’s first order

²⁹In 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 6.

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

Table 7: A summary of the comparative static results under the different models. The last column applies to informed workers only.

belief, and thus ‘Signal’ is not considered as an independent variable in Models 5 and 6 in Table 6. 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 2).

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 1). 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 1). 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 of the results

Table 1 summarizes the comparative static results based on Propositions 1, 2, 3, when there is an interior solution to effort $e^* \in (0, 1)$. Following on, Table 7 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 1 and 7 we see that all models explain well the basic prediction of gift exchange, $\frac{\partial e^*}{\partial w} > 0$. From Table 3 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 (i) the wage norm, θ_w ($\frac{\partial e^*}{\partial \theta_w} > 0$), (ii) the effort norm, θ_e ($\frac{\partial e^*}{\partial \theta_e} > 0$) and, (iii) for the informed worker, the signal s ($\frac{\partial e^*}{\partial s} > 0$). By contrast, the CGE predicts that $\frac{\partial e^*}{\partial \theta_w} < 0$ and $\frac{\partial e^*}{\partial \theta_e} = \frac{\partial e^*}{\partial s} = 0$, while the AGE predicts that $\frac{\partial e^*}{\partial \theta_w} < 0$, $\frac{\partial e^*}{\partial \theta_e} < 0$ and $\frac{\partial e^*}{\partial s} < 0$ for the informed worker.

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) $\frac{\partial e^*}{\partial s} > 0$.

From Table 4 and Figure 2 we see that, at the 1% significance level, $\rho(e, \theta_e) > 0$ for 92.7% of the workers while $\rho(e, \theta_e) < 0$ for 1.04% of the workers.

From Section 8.5 recall that $\rho(e, s)$ was significant at the 1% level for 46 out of 52 of the

informed workers in our new experiments. For these 46 workers, average $\rho(e, s) = 0.94$.

From Table 5 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, 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), which uses a different transmission channel. In the Akerlof model, through 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.

In sum, the evidence strongly supports the role of conditional second order beliefs in explaining gift exchange, and we believe that our paper has made a reasonably compelling case for this assertion.

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 (out of 96) uninformed workers and 82 (out of 96) informed workers report being influenced by the signal of the effort norm, θ_e (Q3 in the survey). Only 3.13% of the workers say that they decrease their effort when the signal 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 1).

In choosing their effort, 81 (out of 96) uninformed workers and 80 (out of 96) informed workers report being influenced by the signal of the wage norm, θ_w (Q4 in the survey). Only 4.17% of the uninformed and 11.46% of the informed workers say that they decreased their effort when the signal 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 1).

In choosing their effort, 92 (out of 96) informed workers reported being influenced by the firm’s expectation of effort (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 is above the wage norm, and 10 informed workers decrease effort only if the wage is below

the wage norm. These results are inconsistent with the AGE and CGE models but consistent with the BGE model (see Table 1). 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 the wage exceeds the wage norm.

12 Summary and conclusion

We considered three models of gift exchange, the classical (CGE), the augmented (AGE) and the belief-based (BGE). Only the BGE model is based on psychological game theory and it explicitly uses belief hierarchies (beliefs and beliefs about beliefs). 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.

1. All three models are able to explain the gift-exchange hypothesis, namely, higher wages induce higher effort.
2. Contrary to the evidence, the CGE and the AGE models predict that the wage norm (typical wage level in similar firms) should influence effort negatively. By contrast, the BGE model predicts a positive effect, which is consistent with the evidence.
3. Contrary to the evidence, the CGE model predicts that the effort norm (typical effort level in similar firms) should not influence effort; and the AGE model predicts a negative effect. By contrast, the BGE model predicts a positive effect, which is supported by our evidence.
4. Contrary to the evidence, the CGE model predicts that an informed worker's belief about the firm's expectation of their effort should have no effect and the AGE model predicts a negative effect. The BGE model predicts a positive effect, which is confirmed by the evidence.

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 satisfactory modelling of the emotion of guilt that underpins all the comparative static results in the BGE model. Thus, 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 modelling the underlying belief hierarchies.

Appendix: Proofs

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

$$\frac{\partial W(e; w, \Gamma_j)}{\partial e} = -40e + \beta(w - \theta_w), \quad (12.1)$$

$$\frac{\partial^2 W(e; w, \Gamma_j)}{\partial e^2} = -40 < 0. \quad (12.2)$$

From (12.2), we see that W is a strictly concave function of $e \in [0, 1]$. Hence, a unique maximizer, $e^*(w, \Gamma_j) \in [0, 1]$, exists. If $w \leq \theta_w$ then, from (12.1), 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, $e^*(w, \Gamma_j) = 0$. This establishes part (a). If $w \geq \frac{40}{\beta} + \theta_w$ then, from (12.1), 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, $e^*(w, \Gamma_j) = 1$. This establishes part (c). Now suppose that $\theta_w < w < \frac{40}{\beta} + \theta_w$. From (12.1) 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 $e^*(w, \Gamma_j) \in (0, 1)$. Consequently, $\left[\frac{\partial W(e; w, \Gamma_j)}{\partial e}\right]_{e=e^*(w, \Gamma_j)} = 0$. Solving this equation, using (12.1), we get $e^*(w, \Gamma_j) = \frac{\beta}{40}(w - \theta_w)$. Part (b) then follows. In the case of part (a), $e^*(w, \Gamma_j) = 0$. In the case of part (c), $e^*(w, \Gamma_j) = 1$. From the proof of part (b) $e^*(w, \Gamma_j) = \frac{\beta}{40}(w - \theta_w)$ in that case. In each case $e^*(w, \Gamma_j)$ does not depend on θ_e or s . This establishes part (d). ■

Proof of Proposition 2: From (4.4) we see that V is a continuous function of $e \in [0, 1]$. Furthermore, we get

$$\frac{\partial V(e; w, \Gamma_j)}{\partial e} = -40e + \sigma \left[w - 100 + \int_{w=0}^{100} P_W^1(w|\Gamma_j) dw \right], \quad (12.3)$$

$$\frac{\partial^2 V(e; w, \Gamma_j)}{\partial e^2} = -40 < 0. \quad (12.4)$$

From (12.4), we see that V is a strictly concave function of $e \in [0, 1]$. Hence, a unique maximizer, $e^*(w, \Gamma_j) \in [0, 1]$, exists.

From (4.4) and Definition 1 we see that $V(e; w, \Gamma_N)$ does not depend on s . Hence, neither can $e^*(w, \Gamma_N)$. This establishes part (a).

If $w \leq E(w|\Gamma_j)$ then, from (12.3), we get $\left[\frac{\partial V(e; w, \Gamma_j)}{\partial e}\right]_{e=0} \leq 0$. Since V is strictly concave, we get $\frac{\partial V(e; w, \Gamma_j)}{\partial e} < 0$, for all $e \in (0, 1]$. Hence, $e^*(w, \Gamma_j) = 0$. This establishes part (b).

If $w \geq \frac{40}{\sigma} + E(w|\Gamma_j)$ then, from (12.3), we get $\left[\frac{\partial V(e; w, \Gamma_j)}{\partial e}\right]_{e=1} \geq 0$. Since V is strictly concave, we get $\frac{\partial V(e; w, \Gamma_j)}{\partial e} > 0$, for all $e \in [0, 1)$. Hence, $e^*(w, \Gamma_j) = 1$. This establishes part (d).

Now suppose that $E(w|\Gamma_j) < w < \frac{40}{\sigma} + E(w|\Gamma_j)$. From (12.3) we then get $\left[\frac{\partial V(e; w, \Gamma_j)}{\partial e}\right]_{e=0} > 0$ and $\left[\frac{\partial V(e; w, \Gamma_j)}{\partial e}\right]_{e=1} < 0$. 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 (12.3), we get $e^*(w, \Gamma_j) = \frac{\sigma}{40} \left[w - 100 + \int_{w=0}^{100} P_W^1(w|\Gamma_j) dw \right]$. Part (c) then follows using (4.5). ■

Proof of Lemma 1: 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 $Eu(w, \Gamma_j) = w - 20 \int_{e=0}^{e=1} e^2 dP_W^2(e|\Gamma_j)$, $j = I, N$. Hence, $\max\{Eu(w, \Gamma_j), w \in [0, 100]\} = 100 - 20 \int_{e=0}^{e=1} e^2 dP_W^2(e|\Gamma_j)$, $j = I, N$ and $\min\{Eu(w, \Gamma_j), w \in [0, 100]\} = -20 \int_{e=0}^{e=1} e^2 dP_W^2(e|\Gamma_j)$, $j = I, N$. From these, and Definition 2d, we get $u^E(w, \Gamma_j) = 100\nu - 20\nu \int_{e=0}^{e=1} e^2 dP_W^2(e|\Gamma_j) - 20(1 - \nu) \int_{e=0}^{e=1} e^2 dP_W^2(e|\Gamma_j)$. 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

$$R(e, w, \Gamma_j) = 100(e - \mu)(w - 100\nu). \blacksquare$$

Proof of Proposition 3: From (5.9) we see that U is a continuous function of $e \in [0, 1]$. Furthermore, we get

$$\frac{\partial}{\partial e} U(e, w, \Gamma_j) = -40e + \lambda_R(w - 100\nu) + \lambda_G - (\lambda_G - \lambda_S) P_W^2(e|\Gamma_j), \quad (12.5)$$

$$\frac{\partial^2}{\partial e^2} U(e, w, \Gamma_j) = -40 - (\lambda_G - \lambda_S) \frac{\partial}{\partial e} P_W^2(e|\Gamma_j), \quad (12.6)$$

$$\frac{\partial^2}{\partial e \partial w} U(e, w, \Gamma_j) = \lambda_R - (\lambda_G - \lambda_S) \frac{\partial}{\partial w} P_W^2(e|\Gamma_j) = \lambda_R, \quad (12.7)$$

$$\frac{\partial^2}{\partial e \partial \theta_w} U(e, w, \Gamma_j) = -(\lambda_G - \lambda_S) \frac{\partial}{\partial \theta_w} P_W^2(e|\Gamma_j), \quad (12.8)$$

$$\frac{\partial^2}{\partial e \partial \theta_e} U(e, w, \Gamma_j) = -(\lambda_G - \lambda_S) \frac{\partial}{\partial \theta_e} P_W^2(e|\Gamma_j), \quad (12.9)$$

$$\frac{\partial^2}{\partial e \partial s} U(e, w, \Gamma_I) = -(\lambda_G - \lambda_S) \frac{\partial}{\partial \theta_e} P_W^2(e|\Gamma_I). \quad (12.10)$$

From (5.2), (5.10) and (12.6) 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 (12.5), 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 $w \geq 100\nu + \frac{40}{\lambda_R}$ then, from (5.2) and (12.5), 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 $100\nu - \frac{\lambda_G}{\lambda_R} < w < 100\nu + \frac{40}{\lambda_R}$. From (5.2) and (12.5) 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,

$$\frac{\partial e^*(w, \Gamma_j)}{\partial w} = - \left[\frac{\frac{\partial^2}{\partial e \partial w} U(e, w, \Gamma_j)}{\frac{\partial^2}{\partial e^2} U(e, w, \Gamma_j)} \right]_{e=e^*(w, \Gamma_j)}, \quad (12.11)$$

$$\frac{\partial e^*(w, \Gamma_j)}{\partial \gamma} = - \left[\frac{\frac{\partial^2}{\partial e \partial \gamma} U(e, w, \Gamma_j)}{\frac{\partial^2}{\partial e^2} U(e, w, \Gamma_j)} \right]_{e=e^*(w, \Gamma_j)}, \quad (12.12)$$

$$\gamma \in \Gamma_j, J \in \{I, N\}. \quad (12.13)$$

Part (c) then follows from (5.2), (5.10), (12.6)-(12.13) and Definition 1. ■

SUPPLEMENTARY MATERIAL: EXPERIMENTAL INSTRUCTIONS

Instructions for all participants

You are now participating in an economics experiment. Please read all the instructions carefully. You may be able to earn some money in the experiment, depending on your decisions and others' decisions. For the entire duration of the experiment, it is not allowed to communicate with other participants in any way. If you have any questions, please raise your hand, and the experimenter will come to your desk as soon as possible.

During the experiment, our currency will not be Chinese Yuan, but tokens. Your total income will first be calculated in tokens. The total amount of tokens that you have accumulated during the experiment will be converted into Chinese Yuan in cash at the end of the experiment at the exchange rate of **1.5 tokens = 1 Yuan**. Additionally, you will receive 20 Yuan, as a participation fee for this experiment.

There are two types of participants in this experiment: firms and workers. Your role (firm or worker) is randomly chosen. The identity of any two participants in the experiments is never revealed to each other. Each firm will be matched with 4 workers.

Note: The workers who are matched with the same firm act completely independently, i.e. they **won't ever** observe the choices of each other, nor will the choices of any one worker affect the material payoffs of any other worker.

The experiment only runs **once**.

Calculation of payoffs of firms and workers

The firm chooses a **wage**, w , for each worker. The wage can be chosen to be any number between 0 and 100 tokens. The same wage applies to all 4 workers who work for the firm.

Any worker who works for the firm chooses a **costly effort level**, e , which can be any number between 0 and 1, after observing the wage paid by the firm. A higher number corresponds to a higher effort and is costlier. You may think of effort as 'quality adjusted units of time' that workers put in for the firm such that an increase in effort means that workers give more quality adjusted units of time to the firm.

The firm's profit from a single worker is calculated as the difference between its revenues and costs:

$$\text{Firm's profit} = 100 \times e - w + 40.$$

The firm earns a revenue equal to $100 \times e$, and pays a cost w from hiring each worker. Thus, the firm's profit is increased by a higher effort from the worker but is decreased by a higher wage.

2. The worker's material payoff is calculated as the difference between the wage rate w paid by the firm and the cost of effort to the worker $c(e) = 20e^2$ (recall that effort, e , is chosen by the worker to be any number between 0 and 1, and higher effort is costlier).

$$\text{Worker's material payoff} = w - 20e^2.$$

Thus, an increase in the wage, w , increases the worker's material payoff, while an increase in the worker's effort, e , reduces the worker's material payoff.

The following table shows some examples of the cost of effort when $c(e) = 20e^2$.

e	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
c(e)	0	0.05	0.2	0.45	0.8	1.25	1.8	2.45	3.2	4.05	5.0
e	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	
c(e)	6.05	7.2	8.45	9.8	11.25	12.8	14.45	16.2	18.05	20	

Thus, the cost of effort to the worker is increasing at an increasing rate.

At the end of the experiment, if the total amount of tokens that you have accumulated during the experiment is negative, then this negative amount will be taken from your participation fee (20 Yuan) to calculate your total income.

Control questions

The numbers in these questions are purely hypothetical and are used to test your understanding of the experiment.

1. A firm pays a wage $w = 10$ tokens to a worker. The worker exerts the effort $e = 0.2$ after observing the wage. Fill in the given space, the (1) profits of the firm _____, and the (2) material payoffs of the worker _____.

2. A firm pays a wage $w = 30$ tokens to a worker. The worker exerts the effort $e = 0.7$. Fill in the given space, the (1) profits of the firm _____, and the (2) material payoffs of the worker _____.

INSTRUCTIONS FOR THE FIRM

This is Stage 1 of the experiment.

The typical effort level exerted by workers in similar firms is $e = 0.4$.

The typical wage paid by firms in similar firms is $w = 40$ tokens.

You, the firm, are matched with 4 workers. Each worker works on an independent project and does not observe the identity or effort choice of any other worker. The effort chosen by any worker is irrelevant for the revenues or costs of the project arising from any other worker. You must choose an identical wage, w , per worker that applies to all 4 workers. Your chosen wage w , will be observed by all 4 workers, who will then choose their effort levels, independently of each other.

At the end of the experiment, you will be randomly matched with 1 of the 4 workers and your profits will be determined ONLY by the wage w that you have chosen and the effort level e that the randomly selected worker has chosen. The effort levels chosen by the remaining 3 workers have no effect on your profits.

You are **completely free** to choose any wage rate w between 0 and 100 tokens.

Your choice of w and the single randomly chosen worker's choice of effort (out of your 4 workers) determine your final profit level.

Your profit from this single worker is given by $(100 - w) \times e$ [w is your chosen wage and e is the chosen effort of the randomly selected worker]. Recall that the randomly selected worker will choose the effort level after observing your choice of a wage w .

Control Questions for the firm

The numbers in these questions are purely hypothetical and are used to test your understanding of the experiment. You are matched with 4 workers.

You choose the wage $w = 50$. When your 4 workers observe this wage, they choose the effort levels $e = 0.6, e = 0.3, e = 0.8$, and $e = 0.5$, independently of each other. At the end of the experiment, one of these workers is randomly chosen. Suppose that this is worker number 3 who has chosen $e = 0.8$.

1. Fill in the given space, the (1) your profits as the firm _____, and the (2) material payoffs of the randomly chosen worker _____.

2. The material payoff of the second worker who has chosen $e = 0.3$ is given by _____.

(End of control questions)

What is your chosen wage between 0 and 100 tokens? _____

What is the effort level that you expect any of your workers to choose between 0 and 1? _____

At the end of the experiment, if the difference between your expectation of your randomly matched worker's effort and his/her actual effort level is strictly less than 0.1, then you will earn an additional prize of 5 Yuan.

End of Experiment

INSTRUCTIONS FOR THE UNINFORMED WORKER

This is Stage 2 of the experiment. In Stage 1, the firm that you are matched with has chosen the wage level, w , to pay to you in Stage 2.

Your task is described in three parts, Part-A, Part-B and Part-C. Please fill in the information requested in the 3 tables, one in each part. You will only receive the instructions for one part at a time. When you complete each part, the experimenter will collect your responses. You will then be given the instructions for the next part. At the end of the experiment, only one part will be randomly chosen to be paid to you.

Part-A

The typical effort level exerted by workers in similar firms is $e = 0.4$.

The typical wage paid by firms in similar firms is $w = 40$ tokens.

For each possible value of the wage, w , chosen by the firm (first row and third row of TABLE 1 below), state your chosen effort level in the second row and fourth row. Your chosen effort must lie between 0 and 1.

TABLE 1

Wage chosen by firm	0	5	10	15	20	25	30	35	40	45	50
Your effort choice											
Wage chosen by firm	55	60	65	70	75	80	85	90	95	100	
Your effort choice											

At the end of the experiment, one of the above 21 columns in TABLE 1 will be chosen. This will correspond to the wage closest to the wage level chosen by the firm in Stage 1. Using this wage, w , and your chosen effort level, e , in TABLE 1, your material payoff is calculated as: $w - 20e^2$.

PART-B

The typical wage paid by firms in similar firms is $w = 40$ tokens.

The firm in Stage 1 has chosen to pay you an actual wage _____ [Here we announce the "actual" wage paid by the firm].

You are given NO INFORMATION about the typical EFFORT level chosen by a worker in similar firms. For each possible typical effort in similar firms and given the wage w that the firm pays you, choose your actual effort level in TABLE 2.

TABLE 2

Typical effort in similar firms	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Your effort choice											
Typical effort in similar firms	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	
Your effort choice											

At the end of the experiment, one of the above 21 columns in TABLE 2 will be randomly chosen. Using the wage, w , chosen by the firm and your chosen effort level, e , in TABLE 2, your material payoff is calculated as: $w - 20e^2$.

PART-C

The firm in Stage 1 has chosen to pay you an actual wage _____ [Here we announce the "actual" wage paid by the firm].

The typical effort level exerted by workers in similar firms is $e = 0.4$.

You are given NO INFORMATION about the typical WAGE level paid by firms in similar firms. For each possible typical wage level paid in similar firms and given the wage w that the firm pays you, choose your effort level (between 0 and 1) in TABLE 3.

TABLE 3

Typical wage in similar firms	0	5	10	15	20	25	30	35	40	45	50
Your effort choice											
Typical wage in similar firms	55	60	65	70	75	80	85	90	95	100	
Your effort choice											

At the end of the experiment, one of the above 11 columns in TABLE 3 will be randomly chosen. Using the wage, w , chosen by the firm and your chosen effort level, e , in TABLE 3, your material payoff is calculated as: $w - 20e^2$.

End of Experiment

INSTRUCTIONS FOR THE INFORMED WORKER

This is Stage 2 of the experiment.

In Stage 1, the firm that you are matched with has chosen the following two things. (1) The wage level, w , to pay to you in Stage 2. (2) The effort level that it expects from you. Note that the firm is not told that its requested effort level will be conveyed to you.

Your task is described in three parts, Part-A, Part-B and Part-C. Please fill in the information requested in the 3 tables, one in each part. You will only receive the instructions for one part at a time. When you complete each part, the experimenter will collect your responses. You will then be given the instructions for the next part. At the end of the experiment, only one part will be randomly chosen to be paid to you.

Part-A

The typical effort level exerted by workers in similar firms is $e = 0.4$.

The typical wage paid by firms in similar firms is $w = 40$ tokens.

The firm that you are matched with, expects you to exert an effort level _____. [The actual requested effort level to be filled in here].

For each possible value of the wage, w , chosen by the firm (first row and third row of the TABLE 1 below), state your chosen effort level in the second row and fourth row. Your chosen effort must lie between 0 and 1.

TABLE 1

Wage chosen by firm	0	5	10	15	20	25	30	35	40	45	50
Your effort choice											
Wage chosen by firm	55	60	65	70	75	80	85	90	95	100	
Your effort choice											

At the end of the experiment, one of the above 21 columns in TABLE 1 will be chosen. This will correspond to the wage closest to the wage level chosen by the firm in Stage 1. Using this wage, w , and your chosen effort level, e , in TABLE 1, your material payoff is calculated as: $w - 20e^2$.

PART-B

The typical wage paid by firms in similar firms is $w = 40$ tokens.

The firm that you are matched with, expects you to exert an effort level _____. [The actual requested effort level to be filled in here].

The firm in Stage 1 has chosen to pay you an actual wage _____. [Here we announce the "actual" wage paid by the firm].

You are given NO INFORMATION about the typical EFFORT level chosen by a worker in similar firms. For each possible typical effort in similar firms and given the wage w that the firm pays you, choose your actual effort level in TABLE 2.

Typical effort in similar firms	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Your effort choice											
Typical effort in similar firms	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1	
Your effort choice											

TABLE 2

At the end of the experiment, one of the above 21 columns in TABLE 2 will be randomly chosen. Using the wage, w , chosen by the firm and your chosen effort level, e , in TABLE 2, your material payoff is calculated as: $w - 20e^2$.

PART-C

The firm in Stage 1 has chosen to pay you an actual wage _____ [Here we announce the "actual" wage paid by the firm].

The typical effort level exerted by workers in similar firms is $e = 0.4$.

The firm that you are matched with, expects you to exert an effort level _____ [The actual requested effort level to be filled in here].

You are given NO INFORMATION about the typical WAGE level paid by firms in similar firms. For each possible typical wage level paid in similar firms and given the wage w that the firm pays you, choose your effort level (between 0 and 1) in TABLE 3.

TABLE 3

Typical wage in similar firms	0	5	10	15	20	25	30	35	40	45	50
Your effort choice											
Typical wage in similar firms	55	60	65	70	75	80	85	90	95	100	
Your effort choice											

At the end of the experiment, one of the above 11 columns in TABLE 3 will be randomly chosen. Using the wage, w , chosen by the firm and your chosen effort level, e , in TABLE 3, your material payoff is calculated as: $w - 20e^2$.

End of Experiment

Post-experimental survey for worker-subjects

Please answer the following questions³⁰.

1. Age:

Gender: (female/male)

Field of study:

Degree of study:

³⁰The post-experimental survey for firm-subjects only has the following Question 1 and 2.

Year of study:

2. Have you participated in similar experiments in the past? (Yes/No)

3. Did the revelation of information about the typical effort level exerted by workers in similar firms influence your choice of effort? Tick the relevant choice.

Yes () No ()

If you ticked No, then go to the next question, otherwise answer the rest of the question.

3a. An increase in the typical effort level in similar firms lead you to (tick one of the following options):

Increase your effort ()

Decrease your effort ()

Sometimes increase, sometimes decrease your effort ()

3b. Would you say that you experienced the emotion of shame if you chose an effort level lower than the typical effort level in similar firms?

Yes () No ()

4. Did the revelation of information about the typical wage level paid to workers in similar firms influence your choice of effort? Tick the relevant choice.

Yes () No ()

If you ticked No, then go to the next question, otherwise answer the rest of the question.

4a. An increase in the typical wage level in similar firms lead you to (tick one of the following options):

Increase your effort ()

Decrease your effort ()

Sometimes increase, sometimes decrease your effort ()

4b. Would you say that you experienced the emotion of shame if you chose an effort level lower than the typical effort level in similar firms, if the wage paid to you is higher than the typical wage paid in similar firms?

Yes () No () Don't know ()

5. [This question is only for the informed subjects treatment] Did the revelation of information about the firm's expectation of your effort have an influence in your chosen effort?

Yes () No ()

If you answered Yes, then answer the rest of this question.

5a. When the firm expects a higher effort, tick which of the following you believe apply to your choices:

Always increase your effort ()

Always decrease your effort ()

Increase your effort only if the wage is above the typical wage paid in similar firms ()

Decrease your effort only if the wage is below the typical wage paid in similar firms ()

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., Dufwenberg M. (2007) Guilt in games. *The American economic review.* 97(2): 170-176.
- [6] Battigalli P., Dufwenberg M. (2009) Dynamic psychological games. *Journal of Economic Theory.* 144(1): 1-35.
- [7] Battigalli P., Dufwenberg M. (2020) Belief-Dependent Motivations and Psychological Game Theory. Mimeo.
- [8] Bellemare, C. and Shearer, B. (2009). Gift Giving and Worker Productivity: Evidence from a Firm Level Experiment. *Games and Economic Behavior* 67: 233-244.
- [9] 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.
- [10] Bellemare, C., Sebald, A. and Suetens, S. (2019). Guilt aversion in economics and psychology. *Journal of Economic Psychology* 73: 52-59
- [11] Bicchieri, C. (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*, Cambridge University Press, New York, USA.
- [12] Blau, P. (1964). *Exchange and Power in Social Life*. New York: Wiley.
- [13] Bowles, S. and Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution*. Princeton University Press: Princeton.
- [14] Camerer C. (2003) *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press: Princeton.
- [15] Cartwright, E. (2019). A Survey of Belief-Based Guilt Aversion in Trust and Dictator Games. *Journal of Economic Behavior and Organization* 167: 430-444.
- [16] Charness, G. (2004). Attribution and reciprocity in a simulated labor market: An experimental investigation. *Journal of Labor Economics.* 22 (3), 665–688.
- [17] Charness and Kuhn (2007). Does pay inequality affect worker effort? Experimental evidence. *Journal of Labor Economics*, 25(4), 693–723.
- [18] 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>

- [19] Dhami, S. (2016) Foundations of behavioral economic analysis. Oxford University Press: Oxford.
- [20] Dhami, S. (2019) Foundations of behavioral economic analysis: Volume II: Other-Regarding preferences. Oxford University Press: Oxford.
- [21] Dhami, S. (2019) Foundations of behavioral economic analysis: Volume IV: Behavioral Game Theory. Oxford University Press: Oxford.
- [22] Dhami, S., Wei, M., and al-Nowaihi, A. (2019a). Public goods games and psychological utility: Theory and evidence. *Journal of Economic Behavior & Organization*. 167: 361–390.
- [23] Dhami, S., Arshad, J. and al-Nowaihi, A. (2019b) Psychological and Social Motivations in Microfinance Contracts: Theory and Evidence. CESifo Working Papers 7773.
- [24] Di Bartolomeo, G., Dufwenberg, M., Papaa, S., and Passarelli, F. (2018) Promises, expectations & causation. *Games and Economic Behavior* 113: 136-146.
- [25] 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.
- [26] Dufwenberg, M., and Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*. 47(2): 268–98.
- [27] Dufwenberg, M., and Kirchsteiger, G. (2019). Modelling kindness. *Journal of Economic Behavior & Organization*. 167: 228–234.
- [28] Ellingsen, T., Johannesson, M., Tjøtta, S. and Torsvik, G. (2010). Testing Guilt Aversion. *Games and Economic Behavior* 68(1): 95–107.
- [29] 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.
- [30] Englmaier, F., and Leider, S. (2012). Contractual and organizational structure with reciprocal agents. *American Economic Journal: Microeconomics*. 4(2): 146–183.
- [31] Falk, A. (2007). Gift exchange in the field. *Econometrica*. 75(5): 1501–1511.
- [32] Falk, A., and Fischbacher, U. (2006). A Theory of reciprocity. *Games and Economic Behavior*. 54(2): 293–315.
- [33] Fehr, E., Gächter, S., and Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: experimental evidence. *Econometrica*. 65(4): 833–860.
- [34] 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.
- [35] Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? An experimental investigation. *Quarterly Journal of Economics*. 108(2): 437–459.
- [36] Fehr, E., Kirchsteiger, G., and Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*. 42(1): 1–34.
- [37] Fehr, E., and Schmidt, K. M. (1999). A Theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*. 114 (3): 817–68.
- [38] Fehr, E. and Schurtenberger, I. (2018). Normative Foundations of Human Cooperation. *Nature Human Behaviour*, 2(7): 458–468.

- [39] 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.
- [40] 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.
- [41] Gächter, S., Nosenzo, D., and Sefton, M. (2012). The impact of social comparisons on reciprocity. *Scandinavian Journal of Economics* 114(4): 1346–67.
- [42] 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.
- [43] Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*. 1(1): 60–79.
- [44] 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.
- [45] Gouldner, A. W. (1960). The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review* 25(2): 161–78.
- [46] Hauge, K. A. (2016) Generosity and guilt: The role of beliefs and moral standards of others. *Journal of Economic Psychology* 54: 35-43.
- [47] 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.
- [48] Inderst, R., Khalmetski, K., and Ockenfels, A. (2019). Sharing Guilt: How Better Access to Information May Backfire. *Management Science* 65(7): 2947-3448.
- [49] Khalmetski K., Ockenfels A, Werner P. (2015) Surprising gifts: Theory and laboratory evidence. *Journal of Economic Theory*. 159: 163–208.
- [50] 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.
- [51] 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.
- [52] Levine, D. K. (1998) Modeling Altruism and Spitefulness in Experiments. *Review of Economic Dynamics* 1(3): 593–622.
- [53] Lucas, R. E. and Rapping, L. A. (1969). Price Expectations and the Phillips Curve. *The American Economic Review* 59(3): 342–350.
- [54] Malmendier, U., and Schmidt, K. M. (2017) You Owe Me. *American Economic Review*. 107(2): 493–526.
- [55] 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).
- [56] Netzer, N., and Schmutzler, A. (2014). Explaining gift-exchange - The limits of good intentions.” *Journal of the European Economic Association*.

- [57] 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.
- [58] Patel, A., and Smith, A. (2019) Guilt and participation. *Journal of Economic Behavior & Organization* 167: 279-295.
- [59] 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
- [60] Rabin M. (1993) Incorporating fairness into game theory and economics. *The American economic review*. 83(5): 1281–1302.
- [61] Schotter, A., and Trevino, I. (2014). Belief Elicitation in the Laboratory. *Annual Review of Economics* 6: 103-128.