On the Power of Surprising Versus Anticipated Gifts *

Rosario Macera†
Pontificia Universidad Católica de Chile
Business School

Vera L. te Velde ‡
University of Queensland

August 26, 2025

Abstract

We study a neglected aspect of the design of gift exchange field experiments: the surprising nature of the treatment group wage increase. We show that if reciprocal workers have expectations-based reference-dependent preferences and expect to work at the market wage, a surprising gift boosts effort above an anticipated one, possibly increasing profits. However, the power of surprises erodes when workers update beliefs rationally, as unanticipated gifts might lead workers to expect further gifts. This harms effort provision and gift profitability when expectations are unfulfilled. Experimental results from a novel repeated gift exchange experiment confirm that surprising gifts increase effort above expected ones; however, their effectiveness diminishes as expectations adjust. We derive implications for the design of gift-exchange experiments.

JEL Classification: D030, J31, J33, M52
Keywords: Reference-dependent preferences, reciprocity, compensation, gift exchange.

1 Introduction

This paper studies the consequences of a seemingly innocuous feature of the standard gift-exchange field experiment: the unexpected wage increase. Following the pioneering work of Gneezy & List (2006), the baseline design increases the wage of a random group of workers hired for a one-time job to compare their productivity to that of a non-gift control group. If reciprocal concerns drive effort provision, the wage increase triggers higher effort as workers reciprocate the gift Akerlof (1984).

*We thank Constanca Esteves-Sorenson for some of the quotes used in this paper. We also thank Stefano DellaVigna, Amanda Friedenberg, Botond Kőszegi and seminar participants at the Australian Economic Theory Workshop, Western Economic Association International annual meetings, University of Technology Sydney, University of New South Wales, Monash University, and the Behavioral-Economic-Theory Workshop at UC Santa Barbara for helpful comments and suggestions. Rosario Macera especially thanks CONICYT Fondecyt Grant Regular #1190305 and ANID PIA/PUENTE AFB230002 for funding.

†rosario.macera@uc.cl.
‡v.tevelde@uq.edu.au.

To avoid selecting abler workers with higher reservation wages, however, all workers are hired at the market wage, and then the raise is surprisingly granted to the treatment group right before executing the job for the first time.

This paper argues that, even though helpful for avoiding selection effects, surprising workers with a wage increase changes the expectations workers hold about the labor relationship, doing more than simply inspiring reciprocity. We study the consequences of the gift's surprising nature by developing and experimentally testing a model of reference-dependent reciprocity, in which expected outcomes constitute the workers' reference points. This model incorporates social preferences into the framework of expectations-based reference dependence developed by Kőszegi & Rabin (2006, 2007, 2009). [1]

Following the model specification, Section 3 theoretically examines reciprocal effort and gift profitability under assumptions replicating a field experiment. In the field, participants believe they are part of a real-world job, where last-minute wage changes are rare. Consequently, their reference belief is receiving the market wage, making any subsequent increase entirely surprising. Under this assumption, we show that an unexpected gift causes a temporarily larger effort response than an anticipated gift of the same magnitude. Becajse a surprising wage increase is a pleasing departure from the expectation of working at the market wage, reciprocal effort can exceed what standard reciprocal preferences would predict. Whenever the interaction lasts longer and the wage increase is permanent, if workers do not expect further wage increases—as is reasonable in the field—workers acclimate to the higher fixed wage and effort wanes back to baseline levels. These predictions align with empirical evidence showing that gifts are most effective when they are surprising (Gilchrist, Luca & Malhotra (2016), Sliwka & Werner (2017), Sautua (2023)) and that the reciprocal effort in response to surprising wage increases may be transitory (Gneezy & List (2006), Bellemare & Shearer (2009), Sliwka & Werner (2017), Gagnon & Noussair (2020)). We further show that, given these assumptions about workers' wage expectations, the gift's size is a crucial determinant of its profitability. Large gifts, which are common in field experiments (see Esteves-Sorenson (2018)), are not profitable because they are too expensive to reciprocate.

Section 4 relaxes the assumption that workers' beliefs are set as in a gift-exchange field experiment and allows workers to form their expectations rationally. We show that after an initial, fully surprising gift, the power of surprises is not sustainable. Workers with rational expectations will now understand that the firm might grant additional gifts, leading to a (stochastic) expectation of further gifts. These expectations, however, hurt firms' expected profits due to retaliation in the form of lower effort when unfulfilled. Even when the firm manages expectations by granting gifts only occasionally, retaliation always outweighs the benefits of the probabilistic gifts. Surprising

[1] The idea that expectations about an upcoming outcome determine the response to the actual outcome dates back to Vroom (1964). This expectancy theory has been extensively studied and applied in psychology – for a description and meta-analysis of the evidence, see Van Eerde & Thierry (1996) and Ambrose & Kulik (1999). Kőszegi & Rabin developed this idea formally by endogenizing expectation formation in a reference-dependent model with loss aversion as in Kahneman & Tversky (1979, 1991).

gifts, therefore, are more likely to be profitable in shorter interactions, where the benefits of the surprise outweigh the long-term costs resulting from unfulfilled expectations of further wage raises caused by the workers' rational updating of expectations.

Section 5 complements the theoretical analysis by presenting the results of a novel laboratory experiment that uncovers the role of expectations in reciprocal effort. Our design allows us to study the long-term impact of surprising, one-time gifts and permanent gifts (wage raises) without the confounding effects of strategic interactions. Employers in our study choose among seven possible wage schedules that assign wages for all 25 work periods in advance, while workers learn these wages on a period-by-period basis and choose effort levels accordingly, knowing that their choices cannot influence their future wages. We use a scoring rule to directly measure workers' wage expectations before they make their effort choices. We then study how these expectations are related to effort provision in the different wage schedules designed to induce fully surprising and anticipated wage surprises. Our experimental results demonstrate the importance of wage expectations for gift exchange. While we do find sustained gift exchange, the effort response to a surprising wage raise diminishes over time as expectations converge to the new wage.

Section 6 outlines the recommendations of our model for the design of gift exchange field experiments. First, managing expectations is crucial. While surprising gifts can boost effort, they may also create expectations of further gifts, potentially harming future effort provision if unfulfilled. Therefore, experimenters should carefully consider how gifts are announced and delivered, ensuring clear communication about the gift's one-time nature to prevent unintended expectations. Second, the size of the gift matters. Our model suggests that only sufficiently small gifts can be rationally expected to be profitable. However, in experimental settings, the justification for the chosen gift sizes is often lacking. Third, establishing credibility in the labor relationship is essential. Gifts perceived as unprofitable or as out-of-equilibrium play may undermine reciprocity, as workers might interpret them as signals of employer intentions, such as compensating for unpleasant tasks or indicating lax monitoring—intentions that might muddle a behavioral response of reference-dependent reciprocity. Finally, gifts in kind can be beneficial as such gifts are less likely to create expectations of future gifts due to their sporadic nature. These findings underscore the importance of thoughtful design in field experiments to allow reciprocal preferences to manifest if they are indeed relevant for effort provision in the workplace.

This paper offers several contributions to the literature. Our first contribution is to deepen our theoretical understanding of the efficacy of gift exchange. Kranton (1996) shows that gift exchange can persist even if it is inefficient relative to a market interaction, and Dur (2009) shows that gift exchange can arise even with low wages if firms can couple low wages with attention, a valued resource for workers. Benjamin (2015) studies a gift-exchange game in which the worker has fairness concerns to show that the model rationalizes several types of wage rigidities. In contrast, Netzer & Schmutzler (2014) shows that gift exchange does not arise in equilibrium if agents have

intention-based reciprocal preferences, as they cannot interpret profitable wage increases as kind. We extend this literature in two directions. First, we explore gift exchange in repeated interactions. After Akerlof (1982), most of the literature on gift exchange assumes a permanent wage raise over a short period to mimic an above-market fixed wage, drawing attention away from the repeated nature of job-market relationships. A notable exception is Fahn (2023), who incorporates gift exchange into a model of informal relational contracts to show that gift exchange and performance pay are dynamic substitutes over the course of an employee's career. Second, allowing for a repeated interaction allows us to focus on an unexplored angle of gift exchange: the gift's optimal size and periodicity.[2]

Our model also contributes to the literature on the importance of surprises in decision making. In a psychological game-theoretic model, Ruffle (1999) shows that when players use mixed strategies, emotions such as surprise, pride, embarrassment, and disappointment can arise in the equilibrium of a gift-giving game in which agents compare actual actions with their first-order beliefs. Khalmetski, Ockenfels & Werner (2015) model surprising gifts in the context of a dictator game in which agents experience guilt aversion as in Battigalli & Dufwenberg (2007). Unlike these papers, we analyze the profitability of gift exchange by focusing on whether the gift was completely unanticipated or anticipated (completely or stochastically) in one-shot and repeated principal-agent interactions.

Third, our model contributes to the literature on expectations-based reference-dependent preferences by proposing a model of reference-dependent reciprocity applied to the employer-employee relationship. Reference-dependent preferences have been shown to be relevant to economic behavior in a large array of domains, such as financial decisions, insurance, saving, pricing, labor supply, etc. (see DellaVigna (2009), Barberis (2013) and Kőszegi (2014) for reviews). Moreover, laboratory and empirical evidence on expectations as reference points has also recently flourished. Abeler et al. (2011), Gill & Prowse (2012), Marzilli Ericson & Fuster (2011), and Karle, Kirchsteiger & Peitz (2015) present laboratory evidence on the role of expectations as reference points, while Mas (2006), Card & Giuliano (2013), Crawford & Meng (2011), Pope, Price & Wolfers (2018), and Lien, Peng & Zheng (2016) provide empirical evidence. More recently, Gagnon-Bartsch & Bushong (2022) and Bushong & Gagnon-Bartsch (2023) develop and test a model of expectation-based reference-dependent attribution bias.[3] We argue that reference-dependent preferences and social preferences naturally combine to improve our understanding of gift exchange in both field experiments and in

[2]To the best of our knowledge, there are only a handful of papers focusing on the properties of the gift, such as Kube, Maréchal & Puppe (2012), Gilchrist, Luca & Malhotra (2016), and Sliwka & Werner (2017). The literature focusing on preferences or the economic environment, however, is more developed. For instance, Brandts et al. (2010) and Schram, Brandts & Gërxhani (2010) focus on the impact of the structure and size of the market; Charness (2004) study the characteristics of who grants the gift; Hennig-Schmidt, Rockenbach & Sadrieh (2010), Englmaier & Leider (2020) and DellaVigna et al. (2022) focus on the agent's information about the firm's surplus or others in his ability to repay the gift; while Chaudhuri & Sbai (2011), and Hannan, Kagel & Moser (2002) focus on the demographics of the recipient.

[3]For opposing evidence of the role of expectation as the reference point see Heffetz & List (2014) in the context of the endowment effect and Zimmermann (2015) in the context of the timing of information arrival.

real firms.[4]

2 Model Set Up

(1) *Preferences.* A principal (the employer) hires an agent (the worker) to exert effort $e \in \mathbb{R}$ for a fixed wage $w \in \mathbb{R}_+$. The principal's preferences are:

Assumption 1 *(Principal's preferences). The principal is a risk-neutral profit maximizer with no "behavioral" components. Her profit function is $\pi(e, w) = be - w$ where $b > 0$.*

The worker experiences utility from two sources: standard consumption utility from material outcomes (Assumption 2) and reference-dependent utility from comparing actual outcomes with a reference point (Assumption 3).[5] Define $\underline{w}$ and w_h as the market wage (which will also correspond to the worker's reservation wage) and an above-market clearing wage, respectively, where $w_h > \underline{w} > 0$.

Assumption 2 *(Consumption utility in wages and effort).*

(i) *Consumption utility $m(\cdot)$ is linear in the wage w.*
(ii) *Effort costs, given by $c(e)$, are minimized at a baseline level $\underline{e} > 0$, and are convex. For simplicity, specify $c(e) = \frac{\gamma}{2}(e - \underline{e})^2$, $\gamma > 0$.*
(iii) *Altogether, consumption utility is given by*

$$m(e, w) = w - \frac{\gamma}{2}(e - \underline{e})^2. \tag{1}$$

In a model with fixed wages, linear consumption utility in wages, as stated in part (i), is an immaterial simplification. Part (ii) states that the cost of effort is convex (as is standard in principal-agent models) and minimized at $\underline{e}$, capturing the intuition that both retaliation and reciprocation in effort may be more costly than simply doing minimum effort. The quadratic effort cost allows for closed-form solutions, which assist intuition. The assumption that the convex effort-cost function

[4] More generally, our model contributes to the recent literature that incorporates reference points (not necessarily expectations) into gift exchange. Dickson & Fongoni (2019) use a fair wage as the reference to study the impact of loss aversion in the asymmetric responses of the wage-effort relationship. Bejarano, Corgnet & Gómez-Miñambres (2021) show that when reciprocity depends on a reference wage that depends on the economic conditions, stability promotes gift exchange. Building on a belief-based reciprocity framework, Zheng, Wang & Ni (2021) experimentally and structurally examines how providing principals with agents' reciprocity information affects wage offers and effort provision. Chan, Feinberg & White (2024) cite a reference-dependent reciprocity framework to rationalize the finding that retained employees by managers who choose to lay off employees, reciprocate their employers more whenever they are evaluated with an imprecise (versus precise) performance measurement system.

[5] To isolate the role of surprises, we delegate the analysis including a pure reciprocity component in $m(\cdot)$ to Online Appendix E. Including non-reference-dependent social preferences in the utility function would introduce a positive baseline correlation between wages and effort, but it would not qualitatively modify the effect of surprises on reciprocal behavior, and would introduce irrelevant complications in the analysis and results. We describe after each of the main propositions how the inclusion of baseline reciprocity affects them.

has a strictly positive interior minimum is also mathematically immaterial but aids with intuition. Finally, part (iii) assumes that consumption utility is additive, as is standard in principal-agent models.[6]

Assumption 3 describes the second component of workers' preferences: reference-dependent utility, that is, utility relative to her expectations. In our model, the two relevant domains are effort and reciprocity, the latter being the novel component.

Assumption 3 *(Reference-dependent utility in effort and wages).*

(i) *Reference-dependent utility is $\eta\mu(x)$, where $\mu(x)$ is piecewise linear with a slope of 1 for $x \geq 0$ and $\lambda > 1$ for $x < 0$ is the loss aversion parameter. $\eta > 0$ represents the relative weight on consumption versus reference-dependent utility.*

(ii) *Given a reference effort $\tilde{e}$, reference-dependent utility in the effort domain is:*

$$n_e(e|\tilde{e}) = \eta\mu(-c(e) + c(\tilde{e})). \tag{2}$$

(iii) *Given a reference wage $\tilde{w}$ and $\tilde{e}$, reference-dependent utility from reciprocity is:*

$$n_k(w, e|\tilde{e}, \tilde{w}) = \alpha\eta\mu(K(w) - K(\tilde{w}))\mu(\pi(e, w) - \pi(\tilde{e}, \tilde{w})) \tag{3}$$

where $K(w)$ represents the firm's kindness function, which is strictly increasing and strictly concave in w, and is equal to zero at the market wage, $K(\underline{w}) = 0$. $\alpha > 0$ corresponds to a reciprocity parameter.

(iv) *Overall reference-dependent utility is given by $n(e, w|\tilde{e}, \tilde{w}) = n_e(e|\tilde{e}) + n_k(w, e|\tilde{w}, \tilde{e})$.*

The assumption that the reference-dependent utility function is piecewise linear, as described in part (i), is standard in applications of reference-dependent preferences and highlights the role of loss aversion in the predictions.[7] Part (ii) describes reference-dependent utility in effort, which

[6] Reciprocity has been shown to impact many different market interactions (see Malmendier, te Velde & Weber (2014) for a review), and substantial evidence exists to support social preferences in gift exchange as well. For instance, Charness (2000), Charness (2004), Charness & Haruvy (2002), Charness et al. (2012), and Huck, Seltzer & Wallace (2011) all run variants of the gift-exchange game in the lab that indicate that social preferences, and intentions-based social preferences, in particular, are key drivers of gift exchange. Laboratory results are in fact so strong that the gift-exchange game has become a workhorse for investigating related phenomena, such as endogenous formation of long-term trading partnerships (Brown, Falk & Fehr 2004, Brown, Falk & Fehr 2012), wage compression (Gross, Guo & Charness 2015, Charness & Kuhn 2007, Güth et al. 2001, Kocher, Luhan & Sutter 2012), wage rigidity (Fehr & Falk 1999), charitable giving (Koppel & Regner 2014), deferred compensation (Huck, Seltzer & Wallace 2011), group decision making (Kocher & Sutter 2007), and peer effects/social comparison (Abeler et al. 2010, Hennig-Schmidt, Rockenbach & Sadrieh 2010, Clark, Masclet & Villeval 2010, Cohn et al. 2014, Gächter & Thöni 2010, Siang, Requate & Waichman 2011).

[7] See, for instance, applications of the Kőszegi & Rabin model to pricing (Heidhues & Kőszegi 2008), labor supply (Crawford & Meng 2011), effort provision (Abeler et al. 2011), sales (Heidhues & Kőszegi 2014), among others.

follows the standard form in applications of the Kőszegi & Rabin (2006) model.[8]

The novel part, reference-dependent reciprocity, is described in part (iii). It assumes that reference-dependent reciprocity depends on a kindness function $K(w)$ that captures how kind the worker believes the employer is by paying a wage w. This function is strictly increasing and concave, capturing the intuition that the marginal utility of kindness is strictly decreasing but always positive, as usually assumed for positive hedonic feelings. Requiring that the kindness function is zero at the market wage reflects the evidence that workers consider the market wage the fair wage and thus neither kind nor unkind (Kahneman, Knetsch & Thaler (1986)). The worker's reference-dependent reciprocity depends on the departure of this actual kindness $K(w)$ with respect to her expectations, $K(\tilde{w})$. Similar to models of intentions-based social preferences such as Rabin (1993), Levine (1998), Dufwenberg & Kirchsteiger (2004), or Falk & Fischbacher (2006), reference-dependent utility from reciprocity assumes the worker puts positive weight on the employer's profits if the employer has been unexpectedly kind and negative weight if unexpectedly unkind. [9]

Finally, Assumption 4 defines the worker's total utility:

Assumption 4 *(Total Utility). Total utility is the sum of consumption and reference-dependent utility,* $U(e, w|\tilde{e}, \tilde{w}) = m(e, w) + n(e, w|\tilde{e}, \tilde{w})$.

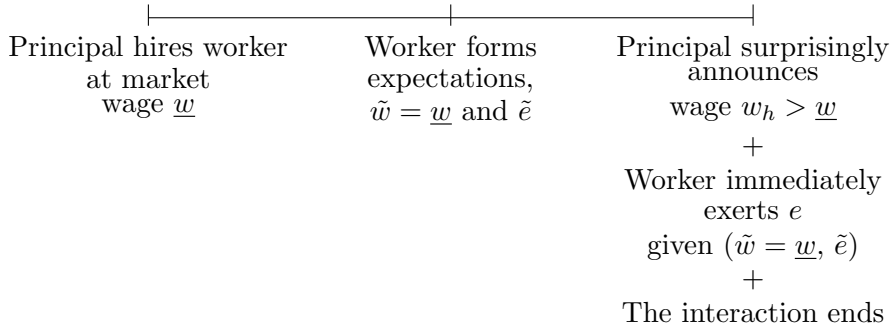
(2) *Reference point formation.* To complete the agent's preferences, we specify how the agent forms her wage and effort expectations. Section 3 replicates a standard gift-exchange field experiment and assumes the worker holds the certain expectation of being paid the wage he was hired at.[10] In Section 4, we relax this assumption and assume workers update their expectations rationally in response to a surprising gift.

[8] Whenever wages are fixed, reference-dependent utility in the monetary dimension does not affect effort provision and thus, without loss of generality, we omit it from the analysis.

[9] Note that the specification in equation (3) departs from the natural extension of the Kőszegi & Rabin (2006) framework, $\alpha\eta\mu(K(w)(be - w) - K(\tilde{w})(b\tilde{e} - \tilde{w}))$. This functional form would make the counterintuitive prediction that a firm that is kind, but not as kind as expected, will *increase* profits as workers try to make up for the loss in overall reciprocity. Another alternative would be $\alpha\eta\mu((K(w) - K(\tilde{w}))(b(e - \tilde{e}) - (w - \tilde{w})))$, but this would predict that workers reward and punish wage deviations to an equal degree, whereas the psychology of loss aversion, along with the empirical evidence in Engelmann & Ortmann (2009) and Kube, Maréchal & Puppe (2013), indicate that workers are more sensitive to wage cuts than wage gains.

[10] Assuming the agent does not foresee a (potentially) profitable gift is the equivalent to assuming that a rational agent has limited knowledge about other players' action sets, the core idea in the unawareness literature (e.g., Modica & Rustichini (1994), Dekel, Lipman & Rustichini (1998*a*), Dekel, Lipman & Rustichini (1998*b*), Aumann (2005)). After Li (2009) showed that the standard state space model could represent this unawareness, the idea that agents can take rational decisions with only partial knowledge of the relevant action spaces has flourished. See Kawamura (2005) for an application to competitive markets, Filiz-Ozbay (2012) for an application to insurance contracts, Auster (2013) for the implications of unawareness for moral hazard, Von Thadden & Zhao (2012) for the market's reaction to heterogeneous levels of awareness in the agents' population, etc. Masatlioglu, Nakajima & Ozbay (2012) study how unawareness affects the revealed preference approach. There is also a related literature on incomplete contracts, which builds on the same idea. See, e.g., Maskin (2002) and Tirole (2009).

Figure 1: Timing and assumptions replicating a gift-exchange field experiment



3 The Power of Fully Surprising Gifts

> *"If your boss walked over your desk and handed you $10,000, would it make you work harder for the rest of the day, or the rest of the year? I think it would!"*
> — Member in Style.Gather.com commenting on Oprah's surprising bonus to her magazine employees in 2010.

This section replicates a gift-exchange field experiment. The principal hires the agent at the market wage $\underline{w}$. Reflecting real-world settings—where wage surprises are rare—and because workers are unaware they're in a study, they expect to earn the market wage at which they were hired, $\tilde{w} = \underline{w}$. Using this wage expectation, the worker forms a belief about how much effort to exert, denoted by $\tilde{e}$. Then the principal grants a surprising wage $w_h > \underline{w}$, and the worker immediately chooses effort e. Due to the short span between the wage increase and the one-time effort provision, beliefs do not update, and workers choose effort given $(\underline{w}, \tilde{e})$. Then the interaction ends. Figure 1 summarizes this timing and assumptions.

We start by defining how effort expectations are formed. Following the framework by Kőszegi & Rabin (2006), we assume that the agent forms their effort expectations based on a *preferred personal equilibrium* or PPE.[11] Definition 1 formally presents this effort plan in this static case with a credible deterministic $\tilde{w}$, following the assumption that, in field experiments, wage expectations are inherited from the workers' experience in real-world jobs.[12]

Definition 1 *(The agent's effort choice with deterministic wage expectations)*
Given a fixed wage expectation $\tilde{w}$, an effort plan $\tilde{e} \in \mathbb{R}$ corresponds to a preferred personal equilibrium (PPE) iff

[11] We focus on pure strategies for the agent, which is without loss of generality until Section 4, where wage expectations will also be formed rationally.

[12] Definition 2 relaxes this assumption and allows workers to form both their effort and wage expectations rationally.

(i) $\tilde{e} \in \underset{e}{\text{argmax}}\, U(e, \tilde{w}|\tilde{e}, \tilde{w})$ *and*

(ii) $\tilde{e} \in \underset{e \in E^*}{\text{argmax}}\, U(e, \tilde{w}|e, \tilde{w})$

where $E^* = \{e \in \mathbb{R} | e$ *solves (i)*$\}$.

Part i) states that, given $\tilde{w}$, the effort plan $\tilde{e}$ must be credible; that is, the worker will want to follow through on it after the formation of the plan affects his reference point. Part ii) states that, if there are multiple credible plans, the worker chooses the one providing the highest utility.

Lemma 1 presents one simple but useful implication of Definition 1 whenever there is no uncertainty in wages:

Lemma 1 *For any fixed wage w, suppose the agent expects $\tilde{w} = w$ with certainty. Then, he forms his effort plans as a consumption utility maximizer, $\tilde{e} = \text{argmax}\ m(e, w) = \underline{e}$. Moreover, if the principal actually grants w, the agent executes his plan, yielding $e = \underline{e}$.*

The intuition behind Lemma 1 is straightforward. Absent uncertainty when forming plans, the reference-dependent utility is zero as expectations are met. Consequently, total utility reduces to consumption utility, and the worker plans to exert minimum effort. If no information arrives, the agent will implement this plan as it is the only credible plan.[13]

3.1 The effort response is stronger for surprising gifts

This section shows that a fully surprising gift leads to greater effort than an anticipated gift. To prove this, notice that, under the assumptions summarized in Figure 1 and the result in Lemma 1, the worker's problem after being surprised with $w_h > \underline{w}$ is to immediately choose an effort level e^* to maximize his utility given his (unrealized) reference point $(\underline{w}, \underline{e})$,

$$e^* \in \underset{e}{\text{argmax}}\, w_h - \frac{\gamma}{2}(e - \underline{e})^2 + \eta\mu\big(-\frac{\gamma}{2}(e - \underline{e})^2\big) + \alpha\eta\mu(K(w_h) - K(\underline{w}))\mu(\pi(w_h, e) - \pi(\underline{w}, \underline{e})) \quad (4)$$

To understand how this re-optimization necessarily leads to an increase in effort relative to an anticipated gift, consider the simpler case where optimal effort occurs at a differentiable point, allowing the first-order condition to be rearranged as

$$e^* = \underline{e} + \frac{\alpha\eta K(w_h)\mu_\pi' b}{\gamma(1 + \eta\lambda)} > \underline{e}, \quad (5)$$

where the inequality holds because $\mu_\pi^{'} = \mu'(\pi(e^*, w_h) - \pi(\underline{e}, \underline{w})) > 0$, which shows that the worker reciprocates the principal's unexpected kindness by exerting more effort than he would have without

[13] Lemma 1 is the application of Proposition 3 in Kőszegi & Rabin (2006) to our setting.

the gift. If the optimization process runs into the kink of the utility function, the worker must compare any interior solutions to the utility he would get from exactly repaying the firm the costs of the gift ($be^* - w_h = b\underline{e} - \underline{w} \Rightarrow e^* = \underline{e} + (w_h - \underline{w})/b$). The full proof in Appendix A explores this case. Proposition 1 presents the full characterization of the effort response, and Online Appendix D analyzes fully surprising wage cuts similarly.

Proposition 1 *A worker who is expecting to receive $\underline{w}$ with certainty but surprisingly receives $w_h > \underline{w}$ and does not have time to update his expectations responds with*

$$e^* = \underline{e} + \begin{cases} \frac{\alpha\eta K(w_h)b}{\gamma(1+\eta\lambda)} & \text{if} \quad w_h < w_g \\ \frac{w_h - \underline{w}}{b} & \text{if} \quad w_g \leq w_h \leq w_l \\ \frac{\alpha\eta\lambda K(w_h)b}{\gamma(1+\eta\lambda)} & \text{if} \quad w_h > w_l \end{cases}$$

where w_g solves $w_g - \underline{w} = \frac{\alpha\eta K(w_g)b^2}{\gamma(1+\eta\lambda)} > 0$ and w_l solves $w - \underline{w_l} = \frac{\alpha\eta\lambda K(w_l)b^2}{\gamma(1+\eta\lambda)} > 0$.

Corollary 1 highlights the main message of the Result 1: fully surprising gifts trigger more effort than if the agent had anticipated the gift.

Corollary 1 *A fully surprising gift $w_h > \underline{w}$ leads to higher effort than a fully anticipated gift of the same magnitude.*

The intuition behind Corollary 1 is simple: whenever the worker anticipates the wage raise and his expectations adapt, such that $\tilde{w} = w_h$, he exerts $\underline{e}$ (see Lemma 1), whichProposition 1 shows is strictly smaller than e^*. Corollary 1 is robust to incorporating baseline reciprocity in an additive form, and Proposition 1 is qualitatively robust to this inclusion. Baseline reciprocity would increase baseline effort when the wage increase is anticipated, but would not affect the additional reciprocity due to the surprise. However, including baseline reciprocity introduces irrelevant complications related to the choice of the baseline wage. For interested readers, Online Appendix E presents the analysis including baseline reciprocity.

3.2 Surprising gifts can be profitable

We now turn to the profitability of a surprising gift in a one-shot interaction, as used in gift-exchange field experiments. Corollary 2 finds the set of gifts that are profitable given the worker's effort response defined in Proposition 1.

Corollary 2 *A fully surprising gift $w_h > \underline{w}$ is profitable only if it is small enough, that is, if $w_h < w_g = \underline{w} + \frac{\alpha\eta K(w_g)b}{\gamma(1+\eta\lambda)}$.*

Intuitively, only gifts that are small enough can be easily reciprocated to keep the firms' profits in the gain domain. Larger gifts, on the contrary, are too expensive in terms of effort costs to reciprocate fully.

Once we have defined the set of profitable gifts, we can look for the profit-maximizing one. Proposition 2 formalizes this optimal gift and describes how it depends on the model parameters.

Proposition 2 *The optimal fully-surprising gift* w_h^*

(i) Exists if $K'(\underline{w}) > \frac{\gamma(1+\eta\lambda)}{\alpha\eta b^2}$*, and*

(ii) Is in the range $\underline{w} < w_h < w_g = \frac{\alpha\eta b^2 K(w)}{\gamma(1+\eta\lambda)}$ *and satisfies* $K'(w_h^*) = \frac{\gamma(1+\eta\lambda)}{\alpha\eta b^2}$*.*

(iii) The range of profitable gifts, the size of the optimal fully surprising gift, the size of the effort response to any gift and to the optimal gift in particular, and the firm's profits after the surprise, are all decreasing in γ *and* λ *and increasing in* α*,* b*, and* η*.*

(iv) Among gifts $w_H > w_h > w_h^*$*,* w_H *is (weakly) the least and* w_h^* *is the most profitable.*

Part (i) shows that a necessary and sufficient condition for the optimal gift to exist is that the kindness function at the market wage is large enough—that is, if the marginal benefit of reciprocating the gift is sufficiently large. Intuitively, if the worker's reciprocal response to the gift is not sufficiently strong, then the gift (even though relatively small) will not be profitable for the firm as the effort response will not compensate for the extra cost.

Part (ii) performs sensitivity analysis on the model parameters. First, a high γ increases the cost of reciprocation, both in terms of consumption and gain-loss utilities. This lowers the revenue-response curve, which in turn reduces the probability that a profitable gift exists, and if it does exist, decreases its size.[14] Second, the optimal gift responds similarly to changes in λ, which measures the impact of losses relative to gains, and γ. A higher λ increases the cost of exerting higher effort to reciprocate the gift, and thus it reduces the optimal gift in the same way that a higher γ does.[15] The parameters α and b both increase the value of reciprocation and thus have a positive impact on the likelihood that an optimal gift exists, and if it exists, on its size. In particular, α has a direct

[14] Experimental evidence supports the prediction that reciprocal effort is decreasing in the marginal cost of effort. Gneezy (2002), using a real-effort task (solving mazes), finds that when difficult "level 5" mazes are used instead of "level-2" mazes, gift exchange is significantly reduced. Interestingly, Gneezy (2002) also finds that the increase in effort when returns to effort are high is *smaller* than when returns are low, but that the former case is profitable, unlike the latter. This is also compatible with a model of reference-dependent preferences, because the worker may be trying to exactly "repay" the employer, and no more. That is, he may be choosing effort to land on the kink in his utility function where he experiences no gains or losses in reciprocity. If he is at that kink (that is $e = \underline{e} + (w_h - \underline{w})/b$), then an increase in b is in fact expected to marginally reduce effort. Engelmann & Ortmann (2009), while not using a real effort task, also find that higher efficiency gains to effort increase both wage offers and effort exerted.

[15] Note that loss aversion only plays a relevant role, from the firm's perspective, through this effort channel: the kinks in the effort response described by Proposition 1 are driven by the switch in μ' from 1 to λ when the firm's profits move from the loss to the gain domain. The only effect of loss aversion in equilibrium is therefore to increase the cost of reciprocation when effort costs are in the loss domain.

impact by increasing the worker's sensitivity to a surprising gift, while b has an indirect impact through improving the returns to effort for the firm.[16] Finally, η behaves differently, since it is a relative weight on any kind of reference-dependent utility, in effort or reciprocity, compared to material (consumption) utility. However, because reciprocity is experienced exclusively as reference-dependent utility, while effort also affects consumption utility, increasing η increases the relative importance of reciprocity compared to effort; therefore, the gain in the reciprocity domain wins out and causes w_h^* to rise.

3.3 The power of gifts wanes over time

"A raise is only a raise for thirty days. After that, it's just somebody's salary."
— Jim Goodnight, CEO of the SAS Institute.

What would happen with the effort provision if the job relationship lasted more than one period? In this section, we preserve the assumption that the firm is able to grant a surprising wage increase without causing the agents to expect further raises— that is, the firm can credibly commit not to raise the wage again. We show that in this case, effort spikes as an immediate response to the wage-raise news, but then wanes back to baseline levels as expectations adapt.

There are infinite periods. In period zero, the principal hires the agent at the market wage $\underline{w}$ for all upcoming periods. Given this wage, the agent forms the belief that he will be paid the market wage for all periods, $\tilde{w}_t = \underline{w}$. By Lemma 1, the agent plans to exert minimum effort in every period, $\tilde{e}_t = \underline{e}$. In period one, the first working period, the principal surprises the agent with a permanent raise to $w_h > \underline{w}$. Following Proposition 1, e_1 rises above the baseline level of $\underline{e}$. The agent, however, then updates his wage and effort expectations for periods two onwards. In particular, he adapts to the new wage and sets $\tilde{w}_t = w_h$ for all $t > 1$.[17] Wage expectations are higher, but still fixed, and so by Lemma 1 he still plans to minimize his cost of effort from period two onwards. In the absence of

[16] The empirical evidence is mixed regarding the prediction that the power of a surprising gift is increasing in the marginal return to effort, b. First of all, Hennig-Schmidt, Rockenbach & Sadrieh (2010) find that clear information about the employer surplus is needed for gift exchange to arise in the lab. Somewhat contrarily, Charness, Frechette & Kagel (2004) show that including a clear payoff table reduces gift exchange, but only among high-effort workers for whom marginal returns to effort are low. In the field with a data entry task, Englmaier & Leider (2020) find that gift exchange arises only whenever workers are informed that their effort leads to a monetary bonus for the principal. More recently, DellaVigna et al. (2022) find no response to the employer's return in their field experiment, whether in the effort domain—where the employer's return is varied through an explicit monetary increase (donor match in the charity experiment)—or in the extra-work domain, where the employer's return is varied by randomly telling subjects that the "extra work is important for us."

[17] The assumption that after a permanent wage increase the agent comes to expect the gift with certainty follows the psychological insight on the hedonic adaptation literature, positing that the hedonic response to a constant stimulus is decreasing in time. Hedonic adaptation was first proposed in psychology by Helson (1964) and applied to economics, among others, by Frederick & Loewenstein (1999) and Frey & Stutzer (2002). After Helson (1964), Brickman & Campbell (1971) coined the term *hedonic treadmill* to apply the concept of adaptation to happiness. See Diener, Lucas & Scollon (2006) for a review.

further surprises, then, effort returns to baseline for all future periods.[18] Corollary 3 summarizes:

Corollary 3 *After a first-period effort increase due to a surprising permanent wage raise $w_t = w_h > \underline{w}$ for all t, effort wanes back to baseline in every period thereafter; that is $e_1 > \underline{e}$ but $e_t = \underline{e}$ for $t > 1$.*

4 Gifts are cursed: The Pitfalls of Anticipated Gifts

> *"[I] Wonder if we will see a Christmas bonus again. I called the Partner Contact Center and they have [stated that] it was a one-time deal because the company made so much money last year."*
>
> — Starbucks employee on the surprising bonus of $250 granted in 2004.

This section examines the consequences of rationality and gift anticipation in labor relationships. Unanticipated gifts from employers lead workers to expect future gifts. We analyze how these rational expectations influence reciprocal effort and examine the firm's optimal strategy for gift-giving.

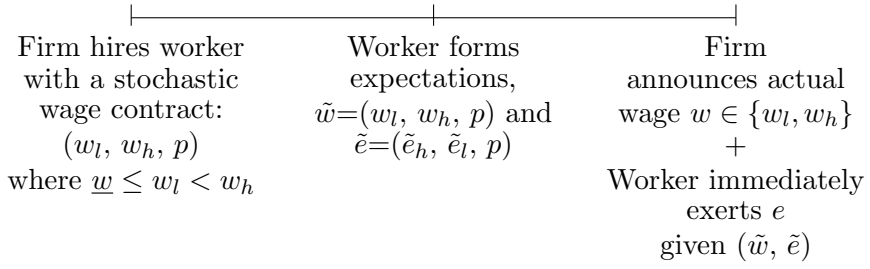
4.1 Better commit to the market wage

> *"I would have loved to give out bonuses, but the thought of people being mad that they didn't get a bonus—a bonus!—when we didn't make enough money to justify it, doesn't make me feel great. If I wanted to be cynical, I could argue that I should have never started giving bonuses in the first place."*
>
> — Jay Goltz, owner of five small businesses in Chicago, writing for the New York Times blog on running small business.

Since our analysis in Section 3 showed that surprising gifts can lead to profitable reciprocity, firms might infer that offering gifts stochastically—hoping to surprise workers at least probabilistically—could be an effective strategy for eliciting higher effort. This section shows, however, that this intuition is incorrect. Once gifts are anticipated, even probabilistically, they become counterproductive: worker retaliation when the gift is not granted outweighs any gains from increased effort when it is.

[18] As discussed in more detail in Online Appendix E, the prediction that effort spikes and then wanes would still hold had we added baseline non-reference-dependent reciprocity to the worker's preferences. Starting from a promised market wage $\underline{w}$, the effort spike would be larger than described in Proposition 1 and effort would wane to a higher level for $t > 1$ as well. Notice that the time required for the effort to wane to baseline levels depends on the speed of adaptation to the reference point, which is treated as instantaneous here. Although more research is needed in this domain, evidence suggests that expectations do adapt fairly quickly but not immediately. In a laboratory experiment varying the time gap that subjects are given between payment information and the execution of a task measuring loss aversion, Song (2016) finds that loss aversion is decreasing in the time gap. Thakral & Tô (2021) provides field evidence of adaptive reference points using data on New York City cab drivers by showing that labor-supply reductions are stronger in response to earnings that accumulate more recently.

Figure 2: Timing and assumptions of the stage game with a stochastic gift.



We derive the stationary equilibria of the infinite principal-agent interaction. In a stationary equilibrium, the principal offers the same (potentially stochastic) compensation scheme in every period, isolating the effect of reference-dependent reciprocity on effort provision.[19] The stage contract is now stochastic: the firm pays a gift wage w_h with probability $p \in (0, 1)$.[20] The baseline non-gift wage is w_l, paid with probability $1 - p$, where $w_h > w_l \geq \underline{w}$. The firm must choose p, w_l, and w_h optimally to maximize profits contingent on the agent's induced expectations.[21,22] Since the agent is rational, the contract (w_l, w_h, p) defines the agent's wage expectations $\tilde{w}=(w_l,\ w_h,\ p)$. Using this stochastic belief, he forms (pure-strategy, for simplicity) contingent effort plans $\tilde{e_l}$ and $\tilde{e_h}$ to execute if w_l or w_h is realized, respectively. We summarize this contingent plan as the tuple $(\tilde{e_l}, \tilde{e_h})$. In each period, the principal announces the actual wage $w \in \{w_h, w_l\}$, the agent immediately exerts effort e maximizing his utility given his stochastic reference point formed by his contingency plan, and the interaction ends. Figure 2 summarizes the stage game timing and assumptions.

Since the gift is now stochastic, Definition 2 formally extends the PPE equilibrium in Definition 1 to consider stochastic reference points and contingent effort plans.

Definition 2 *(The agent's effort choice with stochastic wage expectations)*
Given a stochastic wage expectation $\tilde{w}=(w_l,\ w_h,\ p)$, a contingent effort plan $(\tilde{e_l}, \tilde{e_h})$ constitutes a

[19] Considering non-stationary contracts—e.g., by allowing each period's fixed wage or the probability of getting a gift to be contingent on the game history—confounds the motivation coming from reciprocal preferences with that arising from consumption utility, e.g., career concerns or reputation. We thus focus on the equilibrium where the payment scheme is constant across stages, which is closer to the gift-exchange principle, where no dynamic considerations are needed to motivate workers.

[20] The optimal fixed wage, i.e., $p = 0$ or $p = 1$, when agents have rational expectations is discussed in Section 4.3.

[21] Notice that there are at least two alternative ways of framing this setting. First, the firm offers w_l with certainty and credibly announces a gift $w_g - w_l$, which will be granted stochastically. Second, the firm hires the worker at w_l. Still, the worker knows that the firm is unable to commit to this wage, and therefore forms rational expectations about the probability of a subsequent gift.

[22] Allowing for the base wage to be higher than the market wage, $w_l > \underline{w}$, gives the firm the freedom to compensate workers for overall uncertainty if the ex ante utility of the market wage plus a probabilistic gift does not satisfy the worker's participation constraint.

Preferred Personal Equilibrium (PPE) iff

(i) $\tilde{e}_l \in \underset{e_l}{\text{argmax}}\ EU(e_l, w_l|\tilde{e}, \tilde{w}) = m(e_l, w_l) + EU(n(e_l, w_l|\tilde{e}, \tilde{w}))$

(ii) $\tilde{e}_h \in \underset{e_h}{\text{argmax}}\ EU(e_h, w_h|\tilde{e}, \tilde{w}) = m(e_h, w_h) + EU(n(e_h, w_h|\tilde{e}, \tilde{w}))$ *and*

(iii) $(\tilde{e}_l, \tilde{e}_h) \in \underset{e_l \in E_l^*, e_h \in E_h^*}{\text{argmax}}\ EU(e_l, e_h, \tilde{w}|e_l, e_h, \tilde{w})$

where $E_l^* = \{e \in \mathbb{R} | e$ *solves (i)*$\}$ *and* $E_h^* = \{e \in \mathbb{R} | e$ *solves (ii)*$\}$.

Definition 2 states that choosing effort entails two steps for the worker. First, they must find the set of PE, that is, effort plans that satisfy conditions (i) and (ii). Then, the worker must select the best of these equilibria from an ex ante perspective, that is, they must choose a Preferred PE or PPE (condition (iii)).

In this setting, the firm's problem is to optimally choose w_l, w_h and p in order to maximize profits with the knowledge that the worker will anticipate these choices when setting his wage expectations and choose his effort rationally following Definition 2. The worker additionally only accepts the job if this stochastic contract provides higher expected utility than his reservation wage $\underline{w}$, so the firm is restricted by this participation constraint.[23]

The proof of Proposition 3 in Appendix A formally describes the conditions for an effort plan to be a PPE. Intuitively, when choosing effort, the worker balances their desire to reciprocate, which entails paying a cost to increase effort in the high wage state and also paying a cost to withhold effort in the low wage state, against their desire to avoid effort costs. They attempt to achieve their desired level of reciprocation, measured as the difference in profits between the two states, while keeping effort costs in the two states equal or as close to equal as possible in order to avoid loss aversion when in the higher cost state. This delicate balance crucially depends on p, the likelihood of a gift. For intermediate values of p, the worker may be able to equalize their costs in the high and low wage states; that is, $\tilde{e}_h - \underline{e} = \underline{e} - \tilde{e}_l$. But if p is sufficiently small, the permitted range for $\underline{e} - \tilde{e}_l$ approaches zero, while the permitted range for $\tilde{e}_h - \underline{e}$ remains strictly positive. The opposite happens when p is sufficiently large. We summarize this result in Corollary 4:

Corollary 4 *For a stochastic wage contract* (w_l, w_h, p)*, the worker's PPE* $(\tilde{e}_l, \tilde{e}_h)$ *will satisfy* $\tilde{e}_h - \underline{e}$ *greater (less) than* $\underline{e} - \tilde{e}_l$ *as long as* p *is sufficiently small (large).*

Intuitively, this is because a very low probability p approaches the case of fully surprising gifts: every period that the low wage is paid, workers are only the tiniest bit disappointed at not receiving the high wage, so they choose an effort level close to $\underline{e}$. When the high wage is granted, even though

[23] Notice that the agent has reference-dependent preferences over his outside option too. To keep analysis simple, however, we assume that there is no uncertainty about this payoff and thus only consumption utility is relevant to determine his outside utility.

they understood that this would happen with a small probability, their expectations are greatly exceeded, and they reciprocate with a more substantial increase in effort. When p is very high, the corresponding reasoning makes the reverse true.[24]

Given the worker's optimal reaction to the stochastic contract (w_l, w_h, p) and the firm's problem, we now explore whether firms can improve profits above those they would get by offering the market wage with certainty. Given the characteristics of the worker's PPE, Proposition 3 shows that this is never possible:

Proposition 3 *A stochastic gift exchange arrangement (w_h,w_l,p) is never profitable for the firm compared to credibly committing to pay $\underline{w}$ with certainty.*

This proposition captures the long-run impact of allowing gifts. Once the firm has moved away from committing to a fixed wage, it is left with no option but to commit to an indefinite period of minimizing losses by setting $0 < p \leq 1$. The firm thus is trapped into motivating agents through a clearly suboptimal random gift once it has revealed its inability to commit to paying the market wage. This highlights the importance and difficulty of managing expectations in real-world gift exchange; we discuss this further in Section 6.

Intuitively, loss aversion plays an important role in understanding Proposition 3. Workers expect a gift with some probability, so a realized gift puts the firm's kindness in the gain domain, and an unrealized gift puts it in the loss domain. Workers then punish the losses more than they reward the gains, and this asymmetry is one of the fundamental drivers at work in Proposition 3. No matter how the firm tries to manage expectations—for example by choosing a very small value of p in order for an unrealized gift to induce very little feeling of loss—these small losses over many periods outweigh the gains from the occasional high-wage period.

4.2 Once granted, gifts must be granted forever

This section addresses the long-term consequences of offering gifts. Can firms return to paying the market wage after granting a one-time gift? We show this is not the case. Once granted, the firm reveals its inability to commit to the market wage, and rational expectations dictate that further gifts are probabilistically due. This returns us to the negative impact of expected but unfulfilled gifts, which will force the firm to turn the one-time gift into a permanent wage raise.

We assume there are finite T periods. In the first period, the agent is surprised by a profitable wage increase; consequently, in period two onwards, he rationally infers the probability of another

[24] Adding baseline reciprocity to the stochastic expectations model results in an upward force on effort in any wage state that is kinder than the market wage, or a downward force on effort when the wage is unkind. The qualitative pattern that effort will be higher in the high wage state and that positive reciprocity will exceed (fall short of) negative reciprocity when p is low (high) enough remains true, however, as shown in Online Appendix E.

gift.[25] In period zero, the principal hires the agent at the market wage $\underline{w}$ for all upcoming periods. Given this wage, we assume the agent forms the belief that he will be paid the market wage for all periods, $\tilde{w}_t = \underline{w}$. Using this belief, he forms his effort expectation, $\tilde{e}_t = \underline{e}$. In period one, the first working period, the principal raises the period-one wage to $w_h > \underline{w}$, and the agent immediately exerts effort e_1 to maximize current reference-dependent utility. After this surprising gift, for periods two onwards, the agent updates his wage and effort expectations by inferring the true probability p of receiving the same gift again. In the following periods, the principal announces the actual wage from the set $\{\underline{w}, w_h\}$ and the worker exerts effort, following Definition 2.[26]

In finite interactions, the equilibrium unravels through backwards induction. In the final period, the equilibrium belief p must be zero or one, since we know from the proof of Proposition 3 (see Appendix A) that if the firm chooses an interior probability $p \in (0, 1)$ with which to grant the gift, profits will always be strictly higher in one of the two states. The firm is therefore unwilling to randomize and cannot credibly commit to this probability. The existence of the initial profitable gift means that $p = 0$ is also not credible, so the only remaining option is $p = 1$. Through backwards induction, the same occurs in every earlier period. Proposition 4 shows that the optimal wage level to which the firm can credibly commit is in fact the same as the optimal surprising gift found in Proposition 2.

Proposition 4 *In a finite game, if the firm grants an initial profitable surprising gift, it grants the optimal gift w_h^* described in Proposition 2 in the first period, and then turns the gift into a permanent raise in later periods.*

Similarly to Corollary 5, Proposition 4 highlights how difficult it is for a firm to profitably use surprises in long interactions. In a finite-period game, the firm's initial profit from the surprise must outweigh the losses of paying a higher wage that does not trigger reciprocal effort, due to its anticipated nature, for the remaining periods. Clearly, gift exchange is more likely to be profitable the shorter the interaction. We turn to this issue next.

4.3 Gifts can be profitable in sufficiently short interactions

In this section, we show that when the principal-agent interaction is sufficiently short, if the principal surprises the agent with a gift, it might be profitable even if it generates harmful expectations of

[25] These assumptions replicate anecdotal evidence of real-world firms actually granting surprising discretionary bonuses and allow us to explore the dynamic consequences of such gifts, i.e., workers expecting further gifts. Several cases of surprising discretionary bonuses are found in the popular press. For instance, Oprah Winfrey surprised staff members of *O*, the Oprah Magazine, with a $10,000 check (read the story *here*) and the Grenda family, owners of a Melbourne-based bus company, unexpectedly gave an average bonus of $8,500 to its employees (read the story *here*).

[26] Notice that we do not explicitly assume that the agent updates expectations at the end of periods $t > 1$. As explained in Section 4.1, gift exchange requires a focus on stationary contracts where the size of the gift and the likelihood of getting it are constant across periods. This is because making the gift size or its probability depend on the history of the game would introduce other incentives that would be confounded with reciprocity, such as reputation. Additionally, for this subsection, we restrict $w_l = \underline{w}$ for conceptual simplicity. Of course, if the firm can profitably grant gifts under this restriction, it will certainly be able to with flexible w_l.

further gifts.

We assume there are an infinite number of periods. Proposition 3 shows that once the firm reveals its inability to commit to a wage in period 1, stochastically granting further gifts entails an infinite stream of losses from period 2 onwards relative to what it would have earned by keeping wages—and thus expectations—at $\underline{w}$. Even if the firm tries a fixed wage instead of the stochastic contract, the existence of the initial profitable gift means that promising $\underline{w}$ with $p = 1$ in the future is not credible; any credible fixed wage also entails an infinite stream of losses relative to remaining at the market wage. The principal's decision of whether to take advantage of the profits resulting from an initial surprising gift therefore depends on the comparison between the benefits triggered by the surprise (Proposition 2) and the costs of revealing the inability to commit to paying $\underline{w}$. Because the gains from gift exchange accrue in period 1, and all future periods are losses in expected value, a firm would only choose to use a surprising gift if its discount factor is low, or equivalently, if there is a high probability each period that the work relationship will be discontinued. This proves the following corollary:

Corollary 5 *Gift exchange is profitable in an infinite game only with a sufficiently high period-ending probability.*

5 Experiment

We now report a novel experimental test designed to evaluate the empirical validity of our model's predictions. While our experiment is rooted in the standard laboratory gift-exchange game, examining the relationship between employers' wage choices and workers' effort, it crucially departs by explicitly manipulating the predictability of wage changes and directly measuring workers' wage expectations.

5.1 Experimental details

(1) *The employment game.* Employers and employees interact repeatedly over 25 periods, with each employer being matched with two employees before the start of the interaction; these matches remain fixed throughout the game. Before the first wage period begins, the employer selects a wage profile for each of her workers, determining the fixed wages for all 25 periods in advance.

(2) *Wage profiles (treatments).* Table 1 presents five wage profiles designed to induce surprising or anticipated wage increases of small or large magnitude. The profiles are as follows. After the initial five periods where the wage is fixed at 100 for all treatments (to induce workers to perceive this wage as the market wage), beginning in period 6, two schedules offer a permanent gift that increases wages from 100 to 150 or 200 respectively in all remaining periods (Permanent_{50} and

Permanent$_{100}$), and two schedules provide a single-period gift of either 50 or 100 ECUs (One-time$_{50}$ and One-time$_{100}$).

Table 1: Wage profiles (treatments)

Gift treatment	**Period**		
	1-5	**6**	**7-25**
Permanent$_{100}$	100	200	200
Permanent$_{50}$	100	150	150
One-time$_{100}$	100	200	100
One-time$_{50}$	100	150	100
Control	100	100	100

Notes: Each employer was offered two wage schedules to choose from for each of their workers.

(3) *Employer choice.* After learning the structure of the game, each employer receives an endowment of experimental credit units (ECUs) sufficient to cover any wage profile, even if workers shirk. For each worker, the employer is offered two wage profiles from Table 1 so they can choose how to incentivize that employee. After choosing a wage profile for each worker, employers face no further choices. While employees complete their 25 periods of work, employers use the time to complete a survey for an unrelated research project.

(4) *Employee choices and payoffs.* After completing several comprehension checks to proceed to the 25 work periods, workers are informed of the game structure. In particular, they are informed that the wages for all 25 periods were already chosen by their employer; thus, their wages are independent of their effort choices.[27] Then, workers proceed to the 25 working periods. In each of these periods, the timing is as follows. First, we measure the worker's expected wage for the period; then, the worker is informed of their actual wage; and finally, the worker chooses an effort level between 0 and 100. This effort level is tripled and sent to the employer as profit, after deducting the wage. The worker's payoff for the period is the wage received minus their chosen effort level. After submitting their effort level, workers see their earnings for that period and then advance to the next. These work periods occur without further interaction with the employer.

(5) *Measuring workers' wage expectations.* To ensure our treatments induce wage expectations as desired, we measure expectations directly. In each period, prior to learning the current wage, workers first report their expectations about the wage they will receive. To this end, they provide five probabilities, summing to 100, for each of the potential wage levels to be paid in the current period. Accuracy is incentivized via a quadratic scoring rule, and workers are informed that the

[27] Workers are not informed of the specific wage schedule options that employers choose from, but in the initial instructions, they are informed that the wage in each period has to come from the set {0,50,100,150,200}.

Table 2: Summary statistics by treatment

Treatment	% Female	Baseline effort	% Economics	# Participants
Control	0.43 (0.51)	30.29 (25.66)	0.38 (0.5)	21
One-time$_{100}$	0.62 (0.49)	25 (26.61)	0.33 (0.48)	24
Permanent$_{100}$	0.73 (0.46)	25 (19.94)	0.36 (0.49)	22
One-time$_{50}$	0.35 (0.49)*	26.12 (32.22)	0.35 (0.49)	17
Permanent$_{50}$	0.76 (0.44)	31.41 (22.4)	0.35 (0.49)	17
Overall	0.58 (0.5)	27.37 (25.13)	0.36 (0.48)	101

Notes. The sample includes the overall set of workers. Baseline effort is measured by period-5 effort. Statistically significant differences in means between each treatment and the overall sample are denoted at the 10% (*), 5% (**), and 1% (***) levels.

best way to maximize their earnings is to report their true probabilistic beliefs.[28]

Complete experimental instructions are provided Online Appendix F.

5.2 Results

Nine sessions were conducted in 2019 and 2021 in the University of Queensland Behavioral and Economic Science Laboratory, with 101 subjects participating as workers[29]. Each session lasted approximately half an hour, and participants were compensated with an average of 14 AUD (approximately 9 USD). Table 2 demonstrates balance across treatment groups with respect to the control variables. Furthermore, period-5 effort exhibited no statistically significant differences across treatment groups; we use this as a proxy for baseline effort.

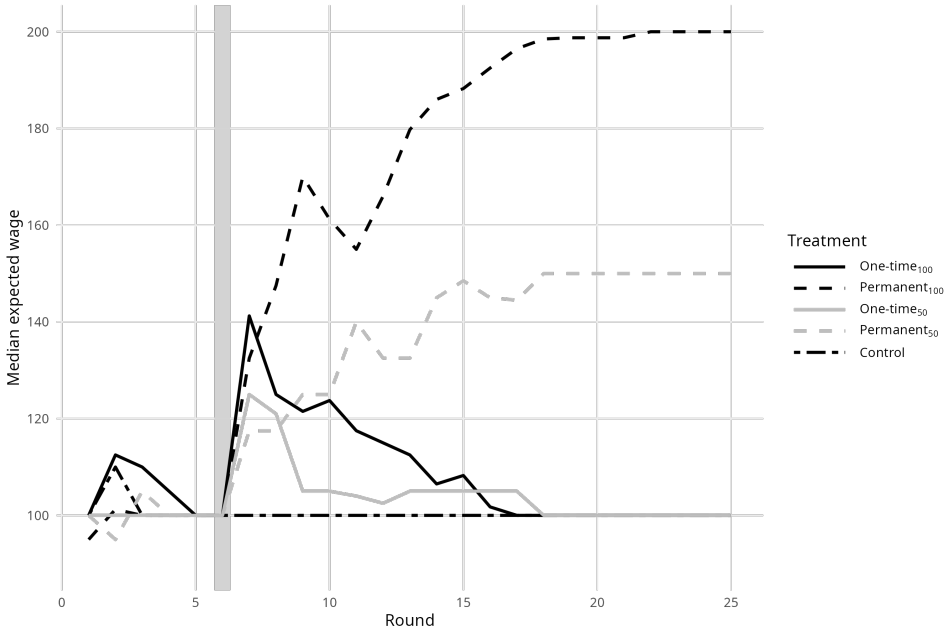
We first demonstrate that the wage profiles (treatments) effectively manipulated expectations. Figure 3 displays the median expected wage across treatments for each period. Initially, the five periods of identical wages successfully established a consistent expectation regarding the baseline wage. Specifically, by period 5, the median expected wage converged to w = 100 across all four treatments and the control group, indicating that the period 6 gift wage created a deviation from this established expectation. Figure 3 also reveals a second key finding: across all wage profiles, expected wages converged with a high degree of accuracy by approximately period 17. That is, the median expected wage aligned with the actual wage offered in each treatment by this time. In contrast, the control group exhibited no deviations from expectations; from period 6 onward, actual wages matched expected wages. These findings are summarized in Result 1.[30]

[28] This follows evidence from Danz, Vesterlund & Wilson (2022) that withholding quantitative information about incentives can lead to higher degrees of truth-telling.

[29] 65 additional workers participated in two additional treatments with randomly repeating gift wages. Because no gifts were profitable in this setting, we omit discussion of these treatments, as we could not use the data to test additional hypotheses. Online Appendix C provides further information about these treatments.

[30] Figure 2 in Online Appendix A shows a more detailed view of the wage expectations. It shows the median beliefs about the probability of each wage by period in each treatment. In all treatments, while people believe there

Figure 3: Median expected wage by wage profile and period.



Result 1 *Expected wages behave as follows:*

(i) ***The Period-6 gift is fully surprising****: In the One-time and Permanent gift treatments, the gift granted in period 6 creates the largest departure from the wage expectation across periods.*

(ii) ***Period 7-16 gifts are partially surprising****: In the One-time and Permanent gift treatments, the gifts granted in period 7-16 create an increasingly smaller departure from the expected wage. By period 17, actual wages match expected wages.*

(iii) ***There are no departures from expectations in the Control group****: Subsequent to period 6, the control group consistently expects to receive the baseline wage of w=100.*

Before analyzing the effort response within thePermanent gift treatments, we first check for any inherent temporal trend in effort. Comparing effort levels across periods necessitates accounting for the potential influence of time itself, such as a general propensity for individuals to adjust their effort choices independently of the treatment. To assess this possibility, we look at the effort trends

is less than a 100% chance of receiving the baseline wage of 100 in period 6, a vast majority of the weight is on that possibility, ensuring that the first gift wage in period 6 is very (if not fully) surprising. Similarly, the weight on the permanent wage (100, 150, or 200) in the Control, Permanent and One-time gift treatments is heavily dominated by the later periods of the game, even if they do recognize a small chance of the other possibilities.

Table 3: Effect of fully surprising and anticipated gifts.

	(1)	(2)	(3)
Full surprise	6.804	6.804	7.412
	(2.642)**	(2.647)**	(4.480)
Partial surprise	3.700	3.700	3.957
	(1.686)**	(1.689)**	(2.542)
Permanent_{100}	9.018	15.836	16.141
	(8.936)	(5.746)***	(6.743)**
Permanent_{100} × Partial surprise			-0.457
			(3.404)
Permanent_{100} × Full surprise			-1.077
			(5.490)
Baseline effort		1.059	1.059
		(0.157)***	(0.157)***
Constant	27.878	24.300	24.128
	(5.743)***	(3.959)***	(4.421)***
N	780	780	780
Controls?	N	Y	Y

Notes. OLS regression of effort levels in the Permanent_{100} and Permanent_{50} treatments, using data from periods 6 to 25. Standard errors are robust and clustered by worker. "Full surprise" is a dummy indicator for period 6, "Partial surprise" for periods 7-17, while the constant captures effort in the non-surprising periods 18-25 in the Permanent_{50} treatment. Controls (gender and economic student status) and baseline effort (effort in period 5) are centered to have mean zero. Statistical significance is indicated at the 10% (*), 5% (**), and 1% (***) levels.

in the Control group, where Result 1, part (iii), shows stable and accurate expectations. In this group, effort remained remarkably stable between periods 6 and 25. Specifically, regressing effort on the period number yields a coefficient of -.0176 ($t = -0.08$), suggesting a minimal decrease of about one-third of a unit of effort across the entire time frame. Consequently, we can confidently attribute any observed changes in effort to deviations from expected wages caused by the wage profiles. (Average effort levels in all periods and all treatments are shown in Figure 1 in Online Appendix A.)

We now turn to the effort response and gift profitability of surprising gifts. Section 3.1 and Section 3.3 showed that workers who have expectations-based reference-dependent preferences exert higher effort whenever gifts are surprising rather than anticipated, but this excess effort wanes as workers' wage expectations adapt to the new wage. To test these predictions, we analyze the effort response of the Permanent gift treatments, in which Result 1 shows that the gift is fully surprising in period 6, partially surprising in periods 7 to 16, and fully anticipated afterwards.

Table 3 supports the theoretical prediction that surprising gifts are effective at eliciting effort, and that this excess effort diminishes as workers' expectations adjust to the new wage. The table displays the OLS regression of effort levels in periods 6 to 25 in the Permanent$_{100}$ and Permanent$_{50}$ treatments, regressed on dummies for period 6 ("Full surprise"), periods 7-17 ("Partial surprise") and the Permanent$_{100}$ treatment.[31] Column (1) shows the raw point estimates. "Full surprise" is large and statistically significant, implying a 25% increase in effort relative to the baseline effort (measured by the constant). "Partial surprise" is also positive and significant, but it is about half the size of the effect of the full surprise. Column (2) incorporates demographic controls and baseline effort (measured as effort in period 5). These control variables are centered such that the constant term represents average effort in the Permanent$_{50}$ treatment during the non-surprising periods (periods 18-25). The results from column (1) remain robust after including these controls. The level effect of the Permanent$_{100}$ treatment also becomes statistically significant, indicating that a larger gift increases (non reference-dependent) effort. Finally, column (3) shows that the large and small wage gifts elicit approximately the same boost in effort: the difference, captured by the Permanent$_{100}$ $\times$ Full surprise coefficient, is small and statistically insignificant. The power reduction when dividing the sample in Column (3) means that the coefficient on Full surprise is no longer statistically significant (p-value 0.102), but the magnitude remains similar. Result 2 summarizes.

Result 2 *The effort response in the Permanent gift treatments is higher in period 6 when the gift is surprising than in periods 7-17 when the surprise is weaker, and lowest from period 18 onwards, once workers' wage expectations have converged to the gift wage.*

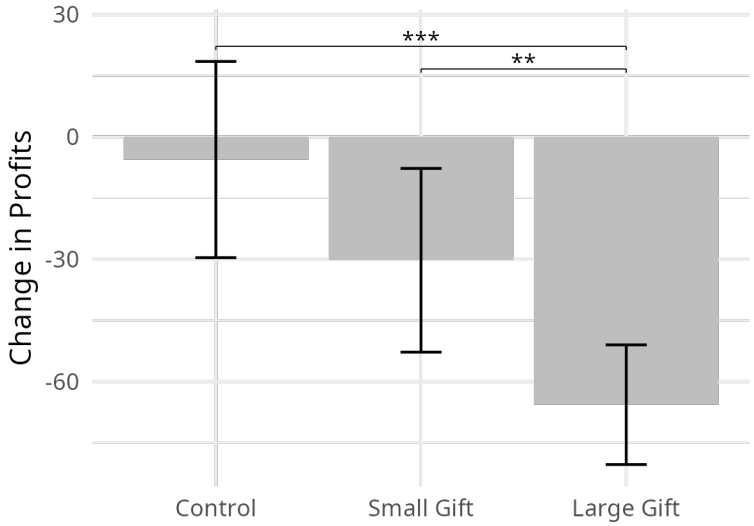
Turning to the profitability of surprising gifts, Figure 4 shows the change in profits between period 5 and 6 depending on the change in wage (i.e., the size of the surprising gift). In the no-gift Control treatment, profits do not noticeably change, as expected. In contrast, the Permanent$_{50}$ and One-time$_{50}$ treatments decrease profits, and the Permanent$_{100}$ and One-time$_{100}$ treatments decrease them significantly further. The wage increase of 50 ECUs reduces profits by an average of 30 ECUs, while that of 100 ECUs reduces profits by an average of 66 ECUs. This aligns with our theoretical prediction that surprising gifts are profitable only if they are sufficiently small, and increasingly unprofitable at larger levels. We have the following result,

Result 3 *Large gifts are more unprofitable than small gifts.*

We now turn to the effort response in the One-time gift treatments. Sections 4.1 and 4.2 showed that the principal is better off committing to the market wage, and that gifts are cursed, as they

[31] The definition of fully anticipated periods as starting in period 17 is not critical to our results; Table 4 in Appendix C shows Column 1 repeated for every possible cutoff from 16 through 25. Additionally, Online Appendix B replicates our results using a continuous measure of the degree of surprise.

Figure 4: Profitability of surprising gifts



Notes. Bars show average changes in profits from baseline (period 5) to immediately after granting a surprising bonus (period 6). No gifts are granted in the Control. One-time$_{50}$ and Permanent$_{50}$ treatments give a surprising small gift of 50 ECUs, and One-time$_{100}$ and Permanent$_{100}$ treatments give a surprising large gift of 100 ECUs. 95% confidence intervals control for demographics and baseline effort. Statistically significant differences in averages are shown at the 10% (*), 5% (**), and 1% (***) levels.

harm effort provision whenever a one-time gift leads to the expectation of further gifts that are not fulfilled. To test how these predictions hold when expectations of further gifts arise after a one-time gift, we analyze the effort response in the One-time gift treatments, where Result 1 shows that between periods 7 and 17, expectations place some weight on the reoccurrence of a gift until the opposite is reinforced and expectations converge back to w=100 by period 17.

Figure 5 shows that after returning to the baseline wage, effort is lower in the One-time$_{50}$ treatment than in the Control treatment, and lower yet in the One-time$_{100}$ treatment, as expected. However, the downward trend is not statistically significant, so while our results suggest that negative reciprocity towards unrepeated bonuses is occurring, we cannot reject the null hypothesis that effort is equal in the three treatments with constant 100 ECU wages in periods 7-25. We summarize in the following result:

Result 4 *Average effort levels are lower following larger one-time surprising bonuses after wages return to baseline levels. However, the differences are not statistically significant.*

Several mechanisms might be limiting the extent of negative reciprocity towards expected but unfulfilled gifts. For instance, there might be aspects of reference-point formation that are not

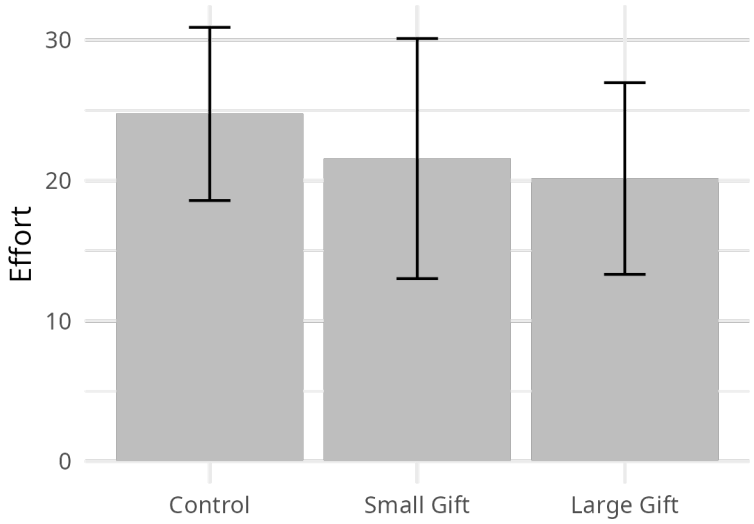


Figure 5: Average effort after returning to baseline wages after one-time surprising gifts. Bars show average effort levels from periods 7 through 25 in the three treatments with constant wages of 100 ECUs during those periods. 95% confidence intervals control for demographics and baseline effort and are clustered at the participant level.

fully captured in our model. For example, positive reciprocity may persist longer than the single period with a one-time gift if recent (rather than current) wages are compared to recent (rather than current) expectations. Alternatively, experimental evidence shows that negative reciprocity is weaker with windfall versus earned money (Danková & Servátka (2015)). In our laboratory setting, this effect could hinder the behavioral response to expected but unfulfilled gifts too. Finally, there is a strand of research showing that behavioral reciprocity is sensitive to perceived intentions (Woods & Servátka (2019)). In our setting, the first five periods—in which all subjects receive the same pre-gift payment—might also hinder subjects' perception of the gift's withdrawal as unkind, and thus mitigate negative reciprocity.

Overall, the experiment evidence supports the view that wage expectations are relevant for effort provision in gift exchange. Unexpected positive changes in wages lead to increased effort, but this effect diminishes as these changes become anticipated. Moreover, larger unexpected positive changes in wages become increasingly unprofitable for employers, revealing the importance of the gift size for reference-dependent reciprocal effort. Finally, there is some indication that one-time unexpected positive changes in wages may lead to lower effort after wages revert to baseline, suggesting potential negative reciprocity, although this effect was not statistically significant in this study.

6 Recommendations for Experimental Designs

(1) *Manage expectations to make the most of gifts.*
Our theoretical predictions emphasize the importance of experimenters explicitly considering the (potentially unintended) wage expectations that their designs may induce. In particular, our model predicts that the most powerful gifts are those that are surprising, even though surprises only trigger temporary effort above baseline. The downside of this surprise, however, is creating the expectation of further gifts, which might harm the firm's future expected profits.

These findings yield two practical recommendations for the design of field experiments, highlighting aspects of gift implementation that have been largely overlooked in the economics literature. First, researchers should exercise caution in how gifts are announced and delivered. Various elements of gift implementation—the announcer's identity, timing, rationale provided, and specific wording used—might influence the wages workers expect to receive in the current and future periods. Experimenters should therefore ensure that workers comprehend the one-time nature of the gift and are informed that no further gifts are forthcoming, while maintaining the external validity and naturalness of this communication.[32]

Second, researchers should carefully consider the duration of the experiment. Extending the study length offers the advantage of capturing the temporal dynamics of gift exchange, particularly as our findings indicate that reference-dependent reciprocity can wane over time. However, longer experiments may increase the likelihood that participants begin to anticipate additional gifts, potentially confounding the results. Experimenters therefore must balance these competing factors when determining the optimal duration of participant engagement. These considerations underscore the necessity of comprehensively documenting every aspect of gift delivery—and any other elements pertinent to expectation formation—in the experimental protocol.

(2) *Be thoughtful about the gift size.*
Beyond the finding that gifts are more powerful when surprising, our model also shows that only sufficiently small gifts can be expected to be profitable. This is directly relevant to experimental designs studying the profitability of gift exchange. Although the definition of 'small' naturally depends on the task and job, current field studies rarely justify the chosen gift size. Consequently, even with similar designs (similar population and tasks), very different gift sizes have been used. For instance,Gneezy & List (2006) granted a 100% raise to data-entry workers, while Englmaier &

[32] A pertinent example of the importance of gift delivery is the announcement made by Disney CEO Bob Iger, who declared that over 125,000 eligible employees would receive a one-time $1,000 cash bonus. The company explicitly and repeatedly emphasized that this bonus was a one-time event resulting from recent tax reforms, thereby clarifying to employees that no further bonuses were due. This approach underscores the significance of clear communication in shaping employees' expectations and responses to unexpected gifts (read Disney's bonus story here). In contrast, the experience of Starbucks in the opening quote of Section 4 illustrates the potential pitfalls of inadequate communication. The company granted an unusual Christmas bonus to its employees without adequately managing expectations regarding the possibility of future gifts. This oversight led to misunderstandings and dissatisfaction among employees, highlighting the critical role of expectation management in the effectiveness of gift-exchange strategies.

Leider (2020) and Kube, Maréchal & Puppe (2013) granted 39% and 20% raises, respectively. See Esteves-Sorenson (2018) and DellaVigna et al. (2022) for a detailed summary of the gift sizes used in field experiments.[33]

The recommendation that experimenters should carefully consider the gift size is consistent with existing laboratory evidence suggesting that information about the profitability of the gift might matter for reciprocal behavior. Hennig-Schmidt, Rockenbach & Sadrieh (2010) show that explicit cost and surplus information enabling exact calculation of an employer's surplus is a necessary condition for reciprocal effort. Charness, Frechette & Kagel (2004) present a similar result by showing that adding a comprehensive payoff table relating wages and effort to workers' and firm's payoffs is necessary to observe reciprocity in a standard gift-exchange game, even though subjects had all the information needed to compute the payoffs on their own. Englmaier & Leider (2020) find that gift exchange increases with the payoff the firm derives from the worker's effort, and this effect is mediated by the worker's personality traits.

(3) *Invest in making the labor relationship credible.*
Our third recommendation is to employ designs that consider the equilibrium play of the firm to achieve identification, which our model indicates is required for external validity. Our theoretical results suggest that unprofitable gift sizes and surprising gifts, if not credibly communicated as one-time events, could be perceived as out-of-equilibrium play by rational workers. In this scenario, it is probable that workers find it difficult to form second-order beliefs to assess the principal's kindness or to forms beliefs more generally. For example, a surprising gift may convey that the employer is trying to compensate for an unexpectedly unpleasant job (as in Bénabou & Tirole (2006)), leading to no or even negative reciprocity, or it could convey that the employer has plenty of resources and so will not be monitoring effort too closely.

Lack of credibility of the gift and labor relationship due to the firm's out-of-equilibrium play could be related to the disparity between the laboratory and field evidence on gift exchange. While laboratory tests have found strong support for gift exchange, field tests have painted a far more dubious picture of reciprocal responses to wage increases. Esteves-Sorenson (2018) and DellaVigna et al. (2022) summarize, showing that in the lab wage-effort elasticities are large, while in the field they are small or non-significant. For instance, Hennig-Schmidt, Rockenbach & Sadrieh (2010) find no gift exchange among students hired to type research abstracts, and Englmaier & Leider (2020) find, if anything, negative gift exchange in their baseline treatment using workers in a data-entry job. Similarly, Kube, Maréchal & Puppe (2013) find no evidence of positive reciprocity to monetary gifts, while Al-Ubaydli et al. (2015) observe minor and insignificant gift exchange among temp workers hired to stuff envelopes. More recently, Esteves-Sorenson (2018) find no significant

[33]This recommendation is in line with Fongoni (2024) who also recommends carefully controlling absolute pay levels when exploring the effects of wage inequality on effort provision. Rojas-Fallas & Williams (2020) provide experimental evidence of asymmetric responses to positive and negative wage inequality.

gift exchange with a data-entry task despite larger than usual gifts, while DellaVigna et al. (2022) also fails to find reciprocal effort in response to monetary gifts in a stuffing-envelope field experiment in the productivity domain, while they do find reciprocal effort in the extra-effort domain.[34]

We believe, however, that the surprising nature and the gift size are not the only design factors that may affect credibility and thus the behavioral response to gifts. The task, recruitment strategy, selection criteria, and work environment are all related to the credibility of the job. Further research is needed to identify how labor relationships that might fail to be credible hurt reciprocal effort, but some results indicate they could be impactful. For instance, Macera & te Velde (2017) show that effort exertion depends on whether the firm is explicit in the wording used to grant the gift, and that effort is expected in return. DellaVigna et al. (2022), however, finds no effect of randomly communicating to workers that their work is of high value to the employer.

(4) *Consider gifts in kind.*
Our theoretical results also suggest the usefulness of gifts in kind when firms have difficulty managing expectations. Gifts in kind have been shown to trigger reciprocal effort to the same (Maréchal & Thöni (2019), DellaVigna et al. (2022)) or even to a greater extent than monetary gifts, even though they are less valued (Kube, Maréchal & Puppe (2012)). According to our model, this could be because of the different expectations induced by different gift types. Unlike monetary gifts, gifts in kind might not create the expectation of further gifts, as their usage is usually more sporadic—for instance, because they are inherently less repeatable (e.g., a company sweater can only reasonably be given every few years) or because they are tied to specific, credibly nonreplicable events, such as hiring, promotions, or holidays.[35] Our theory thus predicts that this property of gifts in kind might be at the heart of the widespread use of this type of incentive in the workplace, and calls for further research on the role of expectations for the optimality of non-monetary gifts.

7 Conclusion

This paper explores the consequences of a seemingly irrelevant aspect of gift-exchange field experiments: the surprising nature of the gift. To avoid selection of abler workers in the baseline design, treated workers are hired at the market wage and then fully surprised with a wage increase right before they work for the first time. A model of expectation-based reference-dependent reciprocity

[34]This disparity has led to a heated debate. Levitt & List (2007) catalogs various reasons why laboratory experiments on social preferences may not extend to the field in general, with specific discussion on the gift exchange literature. Camerer (2015) responds, and on the specific issue of gift exchange, shows that one experiment directly comparing field to lab outcomes, that of List (2006), finds comparable outcomes in the two settings. If a credible gift is necessary to induce reciprocity, then reciprocal effort will naturally be more fragile in the field than in the laboratory, as in the latter, credibility is not an issue as the rules of the game are all explicit and set by the experimenter.

[35]This explanation contrasts with the original explanation in Kube, Maréchal & Puppe (2012) who proposed that gifts in kind are interpreted as kinder than monetary gifts due to the investment of time required from the employer. Within the context of our model, this would represent an amplified kindness function for gifts in kind.

suggests that the gift's surprising nature can trigger a transitorily higher effort response than that of workers in real-world firms where above-market wages are anticipated. We show, however, that this boost in effort might not be exploitable if deviations from the workers' expectations of working at the market wage lead workers to expect additional wage raises, as these expectations will lead to future profit losses when unfulfilled. These results highlight that future field tests of gift exchange should be careful about the unintended effects of the treatment (a surprising wage raise) on beliefs, which can affect the estimation of gift exchange profitability.

Based on these predictions, we derive several recommendations for the design of future field tests of gift exchange, including the explicit manipulation of wage expectations, thoughtful consideration of the gift size, investment in making the labor relationship credible, and evaluation of monetary versus in-kind gifts. These recommendations are, however, not exhaustive. They are based on the surprising aspect of the wage raise and the possibility that departures from expectations mediate reciprocal effort in the field. There exists, however, an array of other issues that are also important for the design of field experiments on gift exchange. We do not touch upon these as they are not related to our model, e.g., asymmetric information beyond that conveyed by the surprise. We thus encourage further theoretical work to improve our understanding of the reach of field experiment designs in gift exchange in other realms.

Finally, we hope our paper offers insights for the design of field experiments beyond the gift-exchange literature. For instance, another example of the use of surprising changes in prices in the economics literature comes from Karlan & Zinman (2008), who exploit a two-stage randomization in a microcredit field experiment to disentangle adverse selection from moral hazard. After borrowers received a randomized interest rate offer in their loan solicitation letter, a second, unexpected randomization at contract signing lowered the interest rate for a subset of borrowers. Because this "surprise" occurred after borrowers had already selected into the loan, any change in repayment behavior can be causally attributed to moral hazard rather than selection. Our theory, however, predicts that the surprising decrease might confound moral hazard with a transitory effect of reciprocity, muddling the conclusions one can get between the relative role of moral hazard and adverse selection. Cai, Luo & Wang (2025) provide another example of using surprise contract terms to isolate moral hazard from adverse selection. In their field experiment, workers were offered a hiring bonus whose amount was only revealed after they accepted the job. Because the bonus was unexpected at the time of decision, any changes in subsequent effort reflect moral hazard, not selection. As in the Karlan & Zinman (2008) case, our theory suggests that the post-hiring bonus's effect might confound moral hazard with a reciprocal response. Our theory, therefore, implies that measuring long-term effects is especially important in designs that rely on surprising price changes for identification, compared to those that do not.

References

Abeler, Johannes, Armin Falk, Lorenz F. Goette, and David Huffman. 2011. "Reference points and effort provision." *American Economic Review*, 101(2): 470–492.

Abeler, Johannes, Steffen Altmann, Sebastian Kube, and Matthias Wibral. 2010. "Gift exchange and workers' fairness concerns: When equality is unfair." *Journal of the European Economic Association*, 8(6): 1299–1324.

Akerlof, George A. 1982. "Labor contracts as partial gift exchange." *Quarterly Journal of Economics*, 97(4): 543–569.

Akerlof, George A. 1984. "Gift exchange and efficiency-wage theory: Four views." *American Economic Review*, 74(2): 79–83.

Al-Ubaydli, Omar, Steffen Andersen, Uri Gneezy, and John A. List. 2015. "Carrots that look like sticks: Toward an understanding of multitasking incentive schemes." *Southern Economic Journal*, 81(3): 538–561.

Ambrose, Maureen L., and Carol T. Kulik. 1999. "Old friends, new faces: Motivation research in the 1990s." *Journal of Management*, 25(3): 231–292.

Aumann, Robert J. 2005. "Musings on information and knowledge." *Econ Journal Watch*, 2(1): 88–96.

Auster, Sarah. 2013. "Asymmetric awareness and moral hazard." *Games and Economic Behavior*, 82: 503–521.

Barberis, Nicholas C. 2013. "The psychology of tail events: Progress and challenges." *American Economic Review*, 103(3): 611–616.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2007. "Guilt in games." *American Economic Review*, 97(2): 170–176.

Bejarano, Hernán, Brice Corgnet, and Joaquín Gómez-Miñambres. 2021. "Economic Stability Promotes Gift-Exchange in the Workplace." *Journal of Economic Behavior & Organization*, 187: 374–398.

Bellemare, Charles, and Bruce Shearer. 2009. "Gift giving and worker productivity: Evidence from a firm-level experiment." *Games and Economic Behavior*, 67(1): 233–244.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and prosocial behavior." *American Economic Review*, 96(5): 1652–1678.

Benjamin, Daniel J. 2015. "A theory of fairness in labor markets." *Japanese Economic Review*, 66(2): 182–225.

Brandts, Jordi, Klarita Gërxhani, Arthur Schram, and Jolanda Ygosse-Battisti. 2010. "Size doesn't matter! Gift exchange in experimental labor markets." *Journal of Economic Behavior & Organization*, 76(3): 544–548.

Brickman, Philip, and Donald T. Campbell. 1971. "Hedonic relativism and planning the good society." In *Adaptation-level theory: A symposium.* , ed. Mortimer Herbert Appley, 287–305. Academic Press.

Brown, Martin, Armin Falk, and Ernst Fehr. 2004. "Relational contracts and the nature of market interactions." *Econometrica*, 72(3): 747–780.

Brown, Martin, Armin Falk, and Ernst Fehr. 2012. "Competition and relational contracts: The role of unemployment as a disciplinary device." *Journal of the European Economic Association*, 10(4): 887–907.

Bushong, Benjamin, and Tristan Gagnon-Bartsch. 2023. "Reference Dependence and Attribution bias: Evidence from Real-Effort Experiments." *American Economic Journal: Microeconomics*, 15(2): 271–308.

Cai, Jing, Sai Luo, and Shing-Yi Wang. 2025. "Money or Monitoring: Evidence on Improving Worker Effort." NBER Working Paper 33977.

Camerer, Colin F. 2015. "The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List." In *Handbook of Experimental Economic Methodology.* , ed. Guillaume R. Frechette and Andrew Schotter, Chapter 14. Oxford University Press.

Card, David, and Laura Giuliano. 2013. "Peer effects and multiple equilibria in the risky behavior of friends." *Review of Economics and Statistics*, 95(4).

Chan, Eric W, Mackenzie K Feinberg, and Brian J White. 2024. "Spread the Pain or Cut Someone Loose? Effects of the Performance Measurement System on Labor Cost Reduction Decisions." *SSRN Working paper*.

Charness, Gary B. 2000. "Responsibility and effort in an experimental labor market." *Journal of Economic Behavior & Organization*, 42(3): 375–384.

Charness, Gary B. 2004. "Attribution and reciprocity in an experimental labor market." *Journal of Labor Economics*, 22(3): 665–688.

Charness, Gary B., and Ernan Haruvy. 2002. "Altruism, equity, and reciprocity in a gift-exchange experiment: An encompassing approach." *Games and Economic Behavior*, 40(2): 203–231.

Charness, Gary B., and Peter J. Kuhn. 2007. "Does pay inequality affect worker effort? Experimental evidence." *Journal of Labor Economics*, 25(4): 693–723.

Charness, Gary B., Guillaume R. Frechette, and John H. Kagel. 2004. "How robust is laboratory gift exchange?" *Experimental Economics*, 7(2): 189–205.

Charness, Gary B., Ramón Cobo-Reyes, Natalia Jiménez, Juan A. Lacomba, and Francisco Lagos. 2012. "The hidden advantage of delegation: Pareto improvements in a gift exchange game." *American Economic Review*, 102(5): 2358–2379.

Chaudhuri, Ananish, and Erwann Sbai. 2011. "Gender differences in trust and reciprocity in repeated gift exchange games." *New Zealand Economic Papers*, 45(1-2): 81–95.

Clark, Andrew E., David Masclet, and Marie-Claire Villeval. 2010. "Effort and comparison income: Experimental and survey evidence." *Industrial and Labor Relations Review*, 63(3).

Cohn, Alain, Ernst Fehr, Benedikt Herrmann, and Frédéric Schneider. 2014. "Social comparison and effort provision: Evidence from a field experiment." *Journal of the European Economic Association*, 12(4): 877–898.

Crawford, Vincent P., and Juanjuan Meng. 2011. "New York City cabdrivers' labor supply revisited: reference-dependence preferences with rational-expectations targets for hours and income." *American Economic Review*, 101(5): 1912–1932.

Danková, Katarína, and Maroš Servátka. 2015. "The house money effect and negative reciprocity." *Journal of Economic Psychology*, 48: 60–71.

Danz, David, Lise Vesterlund, and Alistair J. Wilson. 2022. "Belief Elicitation and Behavioral Incentive Compatibility." *American Economic Review*, 112(9): 2851–2883.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini. 1998*a*. "Recent developments in modeling unforeseen contingencies." *European Economic Review*, 42(3-5): 523–542.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini. 1998*b*. "Standard state-space models preclude unawareness." *Econometrica*, 66(1): 159–173.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature*, 47(2): 315–372.

DellaVigna, Stefano, John A List, Ulrike Malmendier, and Gautam Rao. 2022. "Estimating Social Preferences and Gift Exchange at Work." *American Economic Review*, 112(3): 1038–1074.

Dickson, Alex, and Marco Fongoni. 2019. "Asymmetric Reference-Dependent Reciprocity, Downward Wage Rigidity, and the Employment Contract." *Journal of Economic Behavior & Organization*, 163: 409–429.

Diener, Ed, Richard E. Lucas, and Christie Napa Scollon. 2006. "Beyond the hedonic treadmill: Revising the adaptation theory of well-being." *American Psychologist*, 61(4): 305–314.

Dufwenberg, Martin, and Georg Kirchsteiger. 2004. "A theory of sequential reciprocity." *Games and Economic Behavior*, 47(2): 268–298.

Dur, Robert. 2009. "Gift exchange in the workplace: Money or attention?" *Journal of the European Economic Association*, 7(2-3): 550–560.

Engelmann, Dirk, and Andreas Ortmann. 2009. "The robustness of laboratory gift exchange: A reconsideration." Working Paper.

Englmaier, F, and S Leider. 2020. "Managerial Payoff and Gift-Exchange in the Field." *Review of Industrial Organization*, 56: 259–280.

Esteves-Sorenson, Constanca. 2018. "Gift exchange in the workplace: Addressing the conflicting evidence with a careful test." *Management Science*, 64(9): 4365–4388.

Fahn, Matthias. 2023. "Reciprocity in Dynamic Employment Relationships." *Management Science*, 69(10): 5816–5829.

Falk, Armin, and Urs Fischbacher. 2006. "A theory of reciprocity." *Games and Economic Behavior*, 54(2): 293–315.

Fehr, Ernst, and Armin Falk. 1999. "Wage rigidity in a competitive incomplete contract market." *Journal of Political Economy*, 107(1): 106–134.

Filiz-Ozbay, Emel. 2012. "Incorporating unawareness into contract theory." *Games and Economic Behavior*, 76(1): 181–194.

Fongoni, Marco. 2024. "Does Pay Inequality Affect Worker Effort? An Assessment of Experimental Designs and Evidence." *Journal of Economic Behavior & Organization*, 220: 697–716.

Frederick, Shane, and George F. Loewenstein. 1999. "Hedonic adaptation." In *Well-being: The foundations of hedonic psychology.* , ed. Daniel Kahneman, Ed Diener and Norbert Schwarz, 302–329. New York:Russell Sage Foundation.

Frey, Bruno S., and Alois Stutzer. 2002. "What can economists learn from happiness research?" *Journal of Economic Literature*, 40(2): 402–435.

Gächter, Simon, and Christian Thöni. 2010. "Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis." *Journal of Economic Behavior & Organization*, 76(3): 531–543.

Gagnon-Bartsch, Tristan, and Benjamin Bushong. 2022. "Learning With Misattribution of Reference Dependence." *Journal of Economic Theory*, 203: 105473.

Gagnon, Nickolas, and Charles N Noussair. 2020. "Reciprocity Under Brief And Long-Time Delays." *Economic Inquiry*, 58(3): 1517–1530.

Gilchrist, Duncan S., Michael Luca, and Deepak Malhotra. 2016. "When 3+1 > 4: Gift Structure and Reciprocity in the Field." *Management Science*, 62(9): 2639 – 2650.

Gill, David, and Victoria Prowse. 2012. "A structural analysis of disappointment aversion in a real effort competition." *American Economic Review*, 102(1): 469–503.

Gneezy, Uri. 2002. "Does high wage lead to high profits? An experimental study of reciprocity using real effort." Working Paper.

Gneezy, Uri, and John A. List. 2006. "Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments." *Econometrica*, 74(5): 1365–1384.

Gross, Till, Christopher Guo, and Gary B. Charness. 2015. "Merit pay and wage compression with productivity differences and uncertainty." *Journal of Economic Behavior & Organization*, 117: 233–247.

Güth, Werner, Manfred Königstein, Judit Kovács, and Enikõ Zala-Mezõ. 2001. "Fairness within firms: The case of one principal and multiple agents." *Schmalenbach Business Review*, 53(April): 82–101.

Hannan, R. Lynn, John H. Kagel, and Donald V. Moser. 2002. "Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences, and effort requests on behavior." *Journal of Labor Economics*, 20(4): 923–951.

Heffetz, Ori, and John A. List. 2014. "Is the endowment effect an expectations effect?" *Journal of the European Economic Association*, 12(5): 1396–1422.

Heidhues, Paul, and Botond Kőszegi. 2008. "Competition and price variation when consumers are loss averse." *American Economic Review*, 98(4): 1245–1268.

Heidhues, Paul, and Botond Kőszegi. 2014. "Regular prices and sales." *Theoretical Economics*, 9(1): 217–251.

Helson, Harry. 1964. *Adaptation-level Theory.* Oxford:Harper and Rowe.

Hennig-Schmidt, Heike, Bettina Rockenbach, and Abdolkarim Sadrieh. 2010. "In search of workers' real effort reciprocity — A field and a laboratory experiment." *Journal of the European Economic Association*, 8(4): 817–837.

Huck, Steffen, Andrew J. Seltzer, and Brian Wallace. 2011. "Deferred compensation in multiperiod labor contracts: An experimental test of Lazear's model." *American Economic Review*, 101(2): 819–843.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica*, 47(2): 263–292.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1986. "Fairness as a constraint on profit seeking: Entitlements in the market." *American Economic Review*, 76(4): 728–741.

Karlan, Dean S, and Jonathan Zinman. 2008. "Credit elasticities in less-developed economies: Implications for microfinance." *American Economic Review*, 98(3): 1040–1068.

Karle, Heiko, Georg Kirchsteiger, and Martin Peitz. 2015. "Loss aversion and consumption choice: Theory and experimental evidence." *American Economic Journal: Microeconomics*, 7(2): 101–120.

Kawamura, Enrique. 2005. "Competitive equilibrium with unawareness in economies with production." *Journal of Economic Theory*, 121(2): 167–191.

Khalmetski, Kiryl, Axel Ockenfels, and Peter Werner. 2015. "Surprising gifts: Theory and laboratory evidence." *Journal of Economic Theory*, 159: 163–208.

Kocher, Martin G., and Matthias Sutter. 2007. "Individual versus group behavior and the role of the decision making procerure in gift-exchange experiments." *Empirica*, 34(1): 63– 88.

Kocher, Martin G., Wolfgang J. Luhan, and Matthias Sutter. 2012. "Testing a forgotten aspect of Akerlof 's gift exchange hypothesis: Relational contracts with individual and uniform wages." IZA Discussion Paper No. 6415.

Koppel, Hannes, and Tobias Regner. 2014. "Corporate social responsibility in the work place: Experimental evidence from a gift-exchange game." *Experimental Economics*, 17: 347–370.

Kőszegi, Botond. 2014. "Behavioral contract theory." *Journal of Economic Literature*, 52(4): 1075–1118.

Kőszegi, Botond, and Matthew Rabin. 2006. "A model of reference-dependent preferences." *Quarterly Journal of Economics*, 121(4): 1133–1165.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-dependent risk attitudes." *The American Economic Review*, 97(4): 1047–1073.

Kőszegi, Botond, and Matthew Rabin. 2009. "Reference-dependent consumption plans." *American Economic Review*, 99(3): 909–936.

Kranton, Rachel E. 1996. "Reciprocal exchange: A self-sustaining system." *American Economic Review*, 86(4): 830–851.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. 2012. "The currency of reciprocity: Gift exchange in the workplace." *American Economic Review*, 102(4): 1644–1662.

Kube, Sebastian, Michel André Maréchal, and Clemens Puppe. 2013. "Do wage cuts damage work morale? Evidence from a natural field experiment." *Journal of the European Economic Association*, 11(4): 853–870.

Levine, David K. 1998. "Modeling altruism and spitefulness in experiments." *Review of Economic Dynamics*, 1(3): 593–622.

Levitt, Steven D., and John A. List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic Perspectives*, 21(2): 153–174.

Lien, Jaimie W., Qingqing Peng, and Jie Zheng. 2016. "Major earthquake experience and present-focused expenditures." Working Paper.

Li, Jing. 2009. "Information structures with unawareness." *Journal of Economic Theory*, 144(3): 977–993.

List, John A. 2006. "The behavioralist meets the market: Measuring social preferences and reputation effects in actual transactions." *Journal of Political Economy*, 114(1): 1–37.

Macera, Rosario, and Vera L. te Velde. 2017. "Pay that pays: The sensitivity of workplace incentives to social context." Working Paper.

Malmendier, Ulrike, Vera L. te Velde, and Roberto A. Weber. 2014. "Rethinking reciprocity." *Annual Review of Economics*, 6: 849–874.

Marzilli Ericson, Keith M., and Andreas Fuster. 2011. "Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments." *Quarterly Journal of Economics*, 126(4): 1879–1907.

Maréchal, Michel André, and Christian Thöni. 2019. "Hidden Persuaders: Do Small Gifts Lubricate Business Negotiations?" *Management Science*, 65(8): 3877–3888.

Mas, Alexandre. 2006. "Pay, Reference Points, and Police Performance." *Quarterly Journal of Economics*, 121(3): 783–821.

Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay. 2012. "Revealed attention." *American Economic Review*, 102(5): 2183–2205.

Maskin, Eric. 2002. "On indescribable contingencies and incomplete contracts." *European Economic Review*, 46(4-5): 725–733.

Modica, Salvatore, and Aldo Rustichini. 1994. "Awareness and partitional information structures." *Theory and Decision*, 37(1): 107–124.

Netzer, Nick, and Armin Schmutzler. 2014. "Explaining gift-exchange - The limits of good intentions." *Journal of the European Economic Association*.

Pope, Devin G., Joseph Price, and Justin Wolfers. 2018. "Awareness Reduces Racial Bias." *Management Science*, 64(11): 4988–4995. Publisher: INFORMS.

Rabin, Matthew. 1993. "Incorporating fairness into game theory and economics." *American Economic Review*, 83(5): 1281–1302.

Rojas-Fallas, Jose, and J Forrest Williams. 2020. "Wage Differences Matter: An Experiment of Social Comparison and Effort Provision When Wages Increase or Decrease." *Games*, 11(4): 59.

Ruffle, Bradley J. 1999. "Gift giving with emotions." *Journal of Economic Behavior & Organization*, 39(4): 399–420.

Sautua, Santiago I. 2023. "Disentangling the influences of positive reciprocity and mood on gift exchange at work." *Journal of Behavioral and Experimental Economics*, 102: 101966.

Schram, Arthur, Jordi Brandts, and Klarita Gërxhani. 2010. "Information, bilateral negotiations, and worker recruitment." *European Economic Review*, 54(8): 1035–1058.

Siang, Ch'ng Kean, Till Requate, and Israel Waichman. 2011. "On the role of social wage comparisons in gift-exchange experiments." *Economics Letters*, 112(1): 75–78.

Sliwka, Dirk, and Peter Werner. 2017. "Wage increases and the dynamics of reciprocity." *Journal of Labor Economics*, 35(2): 299–344.

Song, Changcheng. 2016. "An experiment on reference points and expectations." Working Paper.

Thakral, Neil, and Linh T. Tô. 2021. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review*, 111(8): 2417–2443.

Tirole, Jean. 2009. "Cognition and incomplete contracts." *American Economic Review*, 99(1): 265–294.

Tversky, Amos, and Daniel Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quarterly Journal of Economics*, 106(4): 1039–1061.

Van Eerde, Wendelien, and Henk Thierry. 1996. "Vroom's expectancy models and work-related criteria: A meta-analysis." *Journal of Applied Psychology*, 81(5): 575–586.

Von Thadden, Ernst Ludwig, and Xiaojian Zhao. 2012. "Incentives for unaware agents." *Review of Economic Studies*, 79(3): 1151–1174.

Vroom, Victor Harold. 1964. *Work and Motivation.* New York:Wiley.

Woods, Daniel, and Maroš Servátka. 2019. "Nice to you, nicer to me: Does self-serving generosity diminish the reciprocal response?" *Experimental Economics*, 22(2): 506–529.

Zheng, Kaiming, Xiaoyuan Wang, and Debing Ni. 2021. "Reciprocity Information and Wage Personalization." *China Economic Review*, 68: 101645.

Zimmermann, Florian. 2015. "Clumped or piecewise? Evidence on preferences for information." *Management Science*, 61(4): 740–753.

A Appendix: Proofs

Proof of Lemma 1

Following Definition 1, the agent forms his effort plans rationally to maximize $EU(e, \tilde{w}|\tilde{e}, \tilde{w})$, where $\tilde{e} \in \underset{e \in E}{\text{argmax}}\ EU(e, \tilde{w}|e, \tilde{w})$. Because $\mu(K(\tilde{w}) - K(\tilde{w}))$ is independent of $\tilde{e}$, so gain-loss reciprocity $n_k(\tilde{w}, e|\tilde{w}, \tilde{e}) = 0$ for any $\tilde{e}$. Therefore the agent's problem reduces to $\tilde{e} \in \underset{e \in E}{\text{argmax}}\ m(e, \tilde{w}) + n_e(e|\tilde{e})$ $\Rightarrow \tilde{e} = \underline{e}$. If no further information arrives, the agent implements this effort plan as it is the only credible plan and thus the unique PPE.

Proof of Proposition 1

By Lemma 1, since $\tilde{w} = \underline{w}$, the employee's effort plan is $\tilde{e} = \underline{e}$. The possible actions after hearing about the wage surprise are to increase effort to e_g so that profits are in the gain domain, to choose another effort e_l so that profits are in the loss domain, to stick with the plan $\underline{e}$, or to exactly compensate the firm for their profit losses due to the wage increase: $be - w_h = b\underline{e} - \underline{w} \Rightarrow e = (w_h - \underline{w})/b + \underline{e}$.

If the optimization process runs into the kink of the utility function the worker must compare any interior solutions to the utility he would get from exactly repaying the firm the costs of the gift ($be^* - w_h = b\underline{e} - \underline{w} \Rightarrow e^* = \underline{e} + (w_h - \underline{w})/b$). Consider first the case where the gift and the worker's optimal response to it increases profits, i.e., $w_h < b(e^* - \underline{e}) + \underline{w}$. Denote e_g (where the subindex stands for gain) as the optimal effort *given* this assumption. Using equation (5) and the requirement that the firm's profits increase with the gift, we have that e_g and the range of profitable gifts are,

$$e_g = \underline{e} + \frac{\alpha\eta K(w_h)b}{\gamma(1+\eta\lambda)} \qquad \Leftrightarrow \qquad w_h < \frac{\alpha\eta K(w_h)b^2}{\gamma(1+\eta\lambda)} + \underline{w} \tag{6}$$

On the other hand, if profits are in the loss domain, $w_h > b(e - \underline{e}) + \underline{w}$, an analogous calculation shows that the optimal response e_l (where the subindex stands for loss) and the range of unprofitable gifts are,

$$e_l = \underline{e} + \frac{\alpha\eta\lambda K(w_h)b}{\gamma(1+\eta\lambda)} \qquad \Leftrightarrow \qquad w_h > \frac{\alpha\eta\lambda K(w_h)b^2}{\gamma(1+\eta\lambda)} + \underline{w} \tag{7}$$

Figure 6 shows the curves defined by the RHS of these profit constraints, so that the inequalities hold with equality when $w_h = w_g$ and $w_h = w_l$ respectively, so that e_g is a valid local optimum when $w_h < w_g$ and e_l is a valid local optimum when $w_h > w_l$.

At a given wage gift w_h the worker must check whether these local optima exist and compare the one(s) that does to the utility he would get from the kinked point in his utility function, where the first order condition doesn't exist. This happens only where profits are exactly equal

to expectations; note that there is no additional kink in the utility function where effort costs are exactly equal to expectations because effort costs are always in the loss domain when $\tilde{e} = \underline{e}$. Additionally, $\underline{e}$ is never preferred because $U(\underline{e} + \epsilon | \underline{e}, \underline{w}) > U(\underline{e} | \underline{e}, \underline{w})$ for small ϵ.

If the worker were to choose effort at the kink-point in his utility function, this would yield

$$U\left(\frac{w_h - \underline{w}}{b} + \underline{e}, w_h | \underline{e}, \underline{w}\right) = w_h - \frac{\gamma}{2}(1 + \eta\lambda)\left(\frac{w_h - \underline{w}}{b}\right)^2.$$

Likewise, the utilities resulting from e_h and e_l when they are true local optima are respectively

$$U(e_g, w_h | \underline{e}, \underline{w}) = w_h + \frac{(\alpha\eta b K(w_h))^2}{2\gamma(1 + \eta\lambda)} - \alpha\eta K(w_h)(w_h - \underline{w})$$

and

$$U(e_l, w_h | \underline{e}, \underline{w}) = w_h + \frac{(\alpha\eta\lambda b K(w_h))^2}{2\gamma(1 + \eta\lambda)} - \alpha\eta\lambda K(w_h)(w_h - \underline{w}).$$

Comparing e_g to $e = \underline{e} + (w_h - \underline{w})/b$, we find that e_g is preferred when

$$\left(w_h - \underline{w} - \frac{b^2\alpha\eta K(w_h)}{\gamma(1 + \eta\lambda)}\right)^2 > 0$$

which is of course always true, and e_l is similarly always preferred when it exists. Therefore, as shown in Figure 6, e_g is chosen when $w_h < w_g$, e_l is chosen when $w_h > w_l$, and in the region where neither is true, the kink point is chosen.

Extreme or corner cases in which w_h and/or w_g are equal to $\underline{w}$ are straightforward to account for.

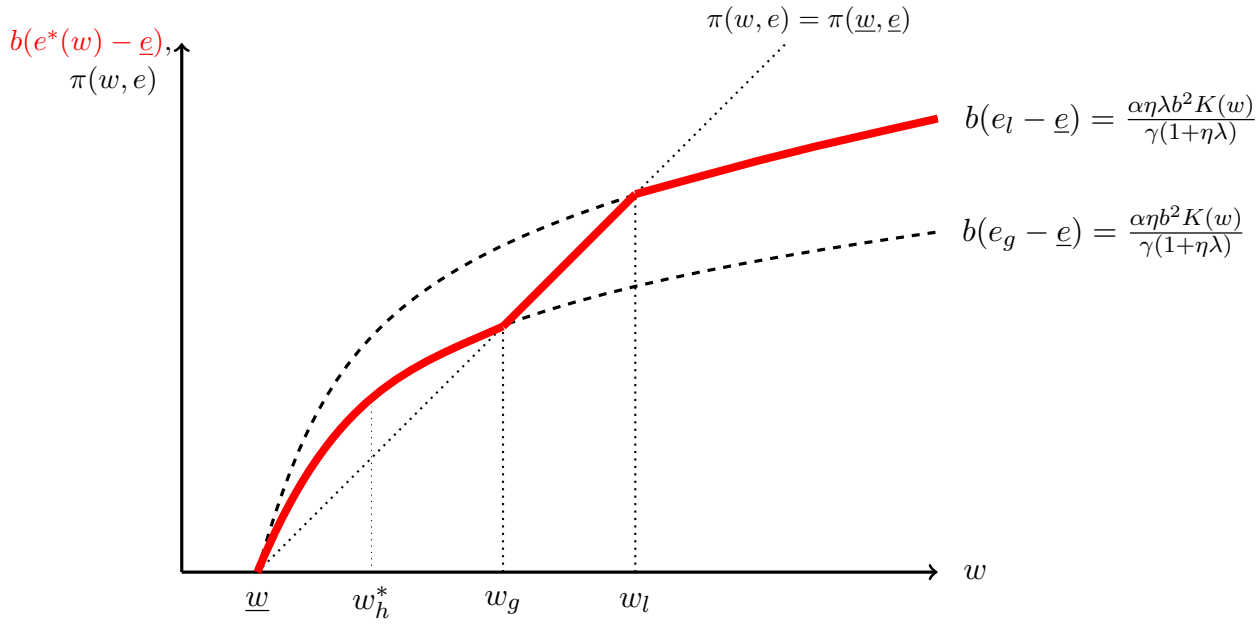
Proof of Corollary 1

Immediate from Proposition 1.

Proof of Corollary 2

To derive the optimal gift, Figure 6 plots the break-even profit function $\pi(w, e) = \pi(\underline{w}, \underline{e})$ and the optimal revenue-response function, $b(e^*(w) - \underline{e})$, as a function of the wage w. First, the break-even profit function shows the effort, and thus revenue, response to a wage w needed for the firm to get the same profits as she would get by granting the market wage. Notice that above this diagonal the firm experiences profits, and below it losses, relative to paying the market wage. Second, the optimal revenue-response function, pictured by a thick red line, represents the revenue the firm gets when granting w given the worker's optimal effort response e^*, described in Proposition 1. To

Figure 6: Revenue-response function in a one-time interaction with profitable gifts.



understand the shape of this function, notice that the lower dotted concave curve, $b(e_g - w)$, shows the revenue increase that occurs when the worker responds with e_g. Since he only does this when profits increase as a result of the gift, he follows this curve as long as it is above the diagonal, or equivalently, if gifts are smaller than the threshold $w_g = \underline{w} + \frac{\alpha\eta K(w_g)b}{\gamma(1+\eta\lambda)}$. Similarly, the upper dotted concave curve, $b(e_l - w)$, shows the revenue increase that occurs when the worker chooses e_l, which he does when this revenue does not compensate for the cost of the gift, decreasing profits. Similarly to the gain-domain case, he follows this curve as long as he is in the loss domain, i.e., below the diagonal line or equivalently, if gifts are higher than the threshold $w_l = \underline{w} + \frac{\alpha\eta\lambda K(w_l)b}{\gamma(1+\eta\lambda)}$. In between these regions, which is a nonzero region due to the assumption that $\lambda > 1$ and the concavity of K, the worker chooses e^* at the kink in his utility function where profits are unchanged relative to expectations.

The position of the optimal revenue-response function in thick red relative to the diagonal break-even profit function shows that gifts are strictly profitable only if they are sufficiently small. When gifts are smaller than a cutoff value w_g, the revenue response function lies above the diagonal line and thus the firm experiences profits relative to paying the market wage. Only gifts that are smaller than w_g are small enough that the worker can easily reciprocate them to keep the firm's profits in the gain domain. Gifts above the cutoff value w_l, however, are too big and thus too expensive, in terms of effort cost, to reciprocate fully. Overall, the profitability of gifts declines monotonically once gifts exceed the optimal level.

Proof of Proposition 2

Continuing from the proof of Corollary 1, and referring to Figure 6, we can see that the range of profitable gifts, with $\underline{w} < w_h < w_g$, exists as long as the revenue response curve shown rises above the diagonal. That is, the slope of this curve at $\underline{w}$ must be greater than 1. This condition is equivalent to $K'(\underline{w}) > \frac{\gamma(1+\eta\lambda)}{\alpha\eta b^2}$ since K is concave, proving the first part of the proposition statement.

Given that a profitable gift exists, the optimal gift is the one that maximizes additional profits, which is the point where the revenue response curve is farthest above the wage cost diagonal. The firm's profits are

$$be_g - w_h = \frac{\alpha\eta b^2 K(w_h)}{\gamma(1+\eta\lambda)}$$

which has an FOC equivalent to the stated implicit definition of w_h^*.

Implicitly differentiating the FOC from proposition 2 gives us

$$\frac{\partial w_h^*}{\partial \gamma} = \frac{1+\eta\lambda}{\alpha\eta b^2 K''(w_h)} < 0$$

$$\frac{\partial w_h^*}{\partial \lambda} = \frac{\gamma\eta}{\alpha\eta b^2 K''(w_h)} < 0$$

$$\frac{\partial w_h^*}{\partial \alpha} = \frac{-\gamma(1+\eta\lambda)}{\alpha^2\eta b^2 K''(w_h)} > 0$$

$$\frac{\partial w_h^*}{\partial b} = \frac{-2\gamma(1+\eta\lambda)}{\alpha\eta b^3 K''(w_h)} > 0$$

$$\frac{\partial w_h^*}{\partial \eta} = \frac{-\gamma}{\alpha\eta^2 b^2 K''(w_h)} > 0$$

The envelope theorem gives us:

$$\frac{\partial \Pi^*(w_h)}{\partial \gamma} = -\frac{\alpha\eta b^2}{\gamma(1+\eta\lambda)} K(w_h^*) < 0$$

$$\frac{\partial \Pi^*(w_h)}{\partial \lambda} = -\frac{\alpha\eta^2 b^2 K(w_h^*)}{\gamma(1+\eta\lambda)^2} < 0$$

$$\frac{\partial \Pi^*(w_h)}{\partial \alpha} = \frac{\eta b^2 K(w_h^*)}{\gamma(1+\eta\lambda)} > 0$$

$$\frac{\partial \Pi^*(w_h)}{\partial b} = \underline{e} + \frac{2\alpha\eta b K(w_h^*)}{\gamma(1+\eta\lambda)} > 0$$

$$\frac{\partial \Pi^*(w_h)}{\partial \eta} = \frac{\alpha b^2 K(w_h^*)}{\gamma(1+\eta\lambda)^2} > 0$$

The size of the effort response to the optimal gift is $e_h - \underline{e} = \frac{\alpha\eta b^2 K(w_h^*)}{\gamma(1+\eta\lambda)}$, which when par-

tially differentiated using the partial derivatives of w_h^* calculated above, similarly yields the stated comparative statics. The response to a particular gift is similar.

Finally, the last part of the proposition follows straightforwardly from the concavity of K.

Proof of Corollary 3

This follows trivially from Proposition 5, which describes the worker's PPE effort response, and is stated and proven in Appendix B below.

Proof of Proposition 3

First, note that Lemma 3, stated and proved in Appendix B below, shows that if a firm in fact wants to offer a stochastic contract, relative to paying the market wage with certainty, workers will never respond with equal costs of effort (i.e. perfectly symmetric effort levels around $\underline{e}$). As in the profit domain, the worker prefers to stay away from the kink in his utility function at which costs of effort are equal in all states, and thus equal to expectations in all states.

With this, and following the derivation of the worker's PPE effort response in the proof of Proposition 5 in Appendix B below, we know that the maximum of the worker's utility function occurs away from any kinks, which leads to the following two constraints on the equilibrium effort plan:

$$\frac{\eta\alpha b\Delta K\tilde{\mu}'_{\pi hl}(1-p)}{\gamma(1+\eta\tilde{\mu}'_{chl}(1-p)+\eta\lambda p)} \leq \tilde{e}_h - \underline{e} \leq \frac{\eta\alpha b\Delta K\tilde{\mu}'_{\pi hl}(1-p)}{\gamma(1+\eta\tilde{\mu}'_{chl}(1-p)+\eta p)}$$

$$\frac{\eta\lambda\alpha b\Delta K\tilde{\mu}'_{\pi lh}p}{\gamma(1+\eta\tilde{\mu}'_{clh}p+\eta\lambda(1-p))} \leq -(\tilde{e}_l - \underline{e}) \leq \frac{\eta\lambda\alpha b\Delta K\tilde{\mu}'_{\pi lh}p}{\gamma(1+\eta\tilde{\mu}'_{clh}p+\eta(1-p))}.$$

We can break these constraints into four cases, with $\pi(\tilde{e}_l, w_l) \lesseqgtr \pi(\tilde{e}_h, w_h)$ and $c(\tilde{e}_h) \lesseqgtr c(\tilde{e}_l)$. The two cases in which $\pi(\tilde{e}_l, w_l) > \pi(\tilde{e}_h, w_h)$ can be dispensed with easily, however: If this were the case, the firm would certainly not wish to offer this contract, as they would earn greater profits by offering $\underline{w}$ for sure than in either wage state in the stochastic contract. The remaining two cases are considered in turn.

Case 1: $c(\tilde{e}_h) > c(\tilde{e}_l)$. In this case, the range of possible values for effort levels above become respectively

$$\frac{\eta\alpha b\Delta K(1-p)}{\gamma(1+\eta\lambda)} \leq \tilde{e}_h - \underline{e} \leq \frac{\eta\alpha b\Delta K(1-p)}{\gamma(1+\eta\lambda(1-p)+\eta p)}$$

$$\frac{\eta\lambda^2\alpha b\Delta Kp}{\gamma(1+\eta p+\eta\lambda(1-p))} \leq -(\tilde{e}_l - \underline{e}) \leq \frac{\eta\lambda^2\alpha b\Delta Kp}{\gamma(1+\eta)}.$$

Will a firm willingly offer a stochastic contract that elicits effort in these ranges? In the best case scenario, the condition that the firm's profits exceed their profits when offering the market wage with certainty becomes

$$\begin{array}{rrcl} & p(b\tilde{e}_h - w_h) + (1-p)(b\tilde{e}_l - w_l) & > & b\underline{e} - \underline{w} \\ \Rightarrow & \frac{\eta\alpha b^2\Delta K(1-p)p}{\gamma(1+\eta\lambda(1-p)+\eta p)} - \frac{\eta\lambda^2\alpha b^2\Delta K(1-p)p}{\gamma(1+\eta p+\eta\lambda(1-p))} - pw_h - (1-p)w_l & > & -\underline{w} \\ \Rightarrow & \frac{\eta\alpha b^2\Delta K(1-p)p}{\gamma}\left(\frac{1-\lambda^2}{1+\eta\lambda(1-p)+\eta p}\right) & > & pw_h + (1-p)w_l - \underline{w} \end{array} \tag{8}$$

Because the RHS and the first term on the LHS are both strictly positive, and because the denominator in the parentheses is positive and the numerator is negative, this inequality never holds, so the firm will never want to offer a contract that elicits these effort levels.

Case 2: $c(\tilde{e}_l) > c(\tilde{e}_h)$. In this case, the range of possible effort values are respectively

$$\frac{\eta\alpha b\Delta K(1-p)}{\gamma(1+\eta(1-p)+\eta\lambda p)} \leq \tilde{e}_h - \underline{e} \leq \frac{\eta\alpha b\Delta K(1-p)}{\gamma(1+\eta)}$$

$$\frac{\eta\lambda^2\alpha b\Delta K p}{\gamma(1+\eta\lambda)} \leq -(\tilde{e}_l - \underline{e}) \leq \frac{\eta\lambda^2\alpha b\Delta K p}{\gamma(1+\eta\lambda p+\eta(1-p))}.$$

The best chance the firm has to meet their participation constraint is then

$$\begin{array}{rrcl} & p(b\tilde{e}_h - w_h) + (1-p)(b\tilde{e}_l - w_l) & > & b\underline{e} - \underline{w} \\ \Rightarrow & \frac{\eta\alpha b^2\Delta K(1-p)p}{\gamma(1+\eta)} - \frac{\eta\lambda^2\alpha b^2\Delta K(1-p)p}{\gamma(1+\eta\lambda)} - pw_h - (1-p)w_l & > & -\underline{w} \\ \Rightarrow & \frac{\eta\alpha b^2\Delta K(1-p)p}{\gamma}\left(\frac{1}{1+\eta} - \frac{\lambda^2}{1+\eta\lambda}\right) & > & pw_h + (1-p)w_l - \underline{w} \end{array} \tag{9}$$

Because the RHS and first term on the LHS are positive, a necessary condition for this to hold is that

$$\begin{array}{rrl} & \frac{1}{1+\eta} & > \frac{\lambda^2}{1+\eta\lambda} \\ \Rightarrow & 0 & > \lambda^2(\eta+1) - \eta\lambda - 1 \end{array}$$

Because the RHS is increasing in λ and $\lambda > 1$, the RHS is actually > 0, a contradiction. So the firm never wishes to offer a contract that elicits these effort levels either.

Proof of Proposition 4

After a surprise gift, which reveals the inability of the employer to commit to a wage, the employee must infer a non-zero probability of a further gift: if they did infer a zero probability, the profitability of the first gift implies that a second fully surprising gift would also be profitable, contradicting the inference that the employer would not want to give a further gift.

As we saw in the proof of Lemma 2, any stochastic contract (w_l, w_h, p) will result in unequal wages in the two wage states, causing the firm to strictly prefer the higher profit wage as long as the game is ending after this interaction so that there is no incentive to maintain the worker's

expectations about future wages. But of course in a rational expectations equilibrium, the worker will anticipate this and form an extreme belief with $p = 0$ or $p = 1$.

Since $p = 0$ would immediately renew the possibility of profitable gift exchange, which the firm would certainly want to take, the only remaining possibility for a rational expectations equilibrium is that workers will expect a gift wage with probability 1. If workers accurately anticipate a certain gift wage, this wage must satisfy the firm's PPE conditions. That is, which wage can the firm plan to give that they will not wish to deviate from? This is equivalent to asking: what is the lowest wage from which there is no profitable fully surprising gift? As in the proof of Proposition 2, no profitable gift exists, starting from a base wage of w, if $K'(w) \leq \frac{\gamma(1+\eta\lambda)}{\alpha\eta b^2}$. Since K is concave, the lowest such wage is the one for which this relationship holds with equality. But the w satisfying this equation is exactly w_h^*, the optimal fully surprising gift.

Is there additionally no profitable deviation in the other direction? That is, if w_h^* is anticipated fully but w with $\underline{w} < w < w_h^*$ is actually paid, will this ever lead to higher profits than fulfilling the expectation of w_h^*? It turns out this is never true. To see this, similarly to the proof of Proposition 1, the worker might reciprocate by either keeping profits equal to expectations (which would not constitute a profitable deviation), or by exerting effort $\underline{e}$ as planned (which is easy to show is never preferred to slightly negatively reciprocating), or by negatively reciprocating in part or in full. The optimal effort levels for the latter two are, as before, given by

$$e_g - \underline{e} = \frac{\alpha\eta\lambda(K(w) - K(w_h^*)b}{\gamma(1+\eta\lambda)}$$

and

$$e_l - \underline{e} = \frac{\alpha\eta\lambda^2(K(w) - K(w_h^*)b}{\gamma(1+\eta\lambda)}.$$

Unlike in Proposition 1 though, since K is concave at the base wage w_h^*, the response of e_g is only a valid optimum, if at all, for small cuts relative to w_h^*. As in Proposition 2, it is never a valid optimum if $\frac{\alpha\eta\lambda K'(w_h^*)b}{\gamma(1+\eta\lambda)} > 1$. But by definition of w_h^*, this quantity is equal to $\lambda > 1$, so it is in fact never an optimal response. The firm can therefore never gain by choosing $\underline{w} < w < w_h^*$, and w_h^* is therefore the firm's PPE.

By backwards induction, the same occurs in every earlier period as well. And so, if the firm chooses the optimal fully surprising gift w_h^*, it must choose the same wage in every further period, turning it into a permanent raise.

Proof of Corollary 5

As described in text.

B Appendix: Additional results

Lemma 2 *Given a stochastic wage contract (w_l, w_h, p), the worker's PPE is never such that $\pi(e_l, w_l) = \pi(e_h, w_h)$.*

Proof of Lemma 2

As shown in the proof of Proposition 3 in Appendix A above, the first order conditions in the high and low wage states respectively are

$$U'(e_h w_h|\tilde{e}, \tilde{w}) = -\gamma(1 + (1-p)\eta\mu'_{chl} + p\eta\mu'_{chh})(e_h - \underline{e}) + (1-p)\alpha\eta b\Delta K\mu'_{\pi hl}$$

and

$$U'(e_l, w_l|\tilde{e}, \tilde{w}) = -\gamma(1 + (1-p)\eta\mu'_{cll} + p\eta\mu'_{clh})(e_l - \underline{e}) - p\alpha\eta\lambda b\Delta K\mu'_{\pi lh}$$

which are defined where $\mu'_{\cdot}$ is defined.

To determine whether an effort level at a kink in the utility function(s) is part of a valid plan, we can consider how the utility function(s) conditional on planning to exert this effort looks and determine whether following through on that plan is at least a local optimum, with U' switching from positive to negative at that kink. We can, in this way, show that planning to exert effort such that $\pi(e_h, w_h) = \pi(e_l, w_l)$ is never a valid plan.

Assume towards contradiction that such a plan exists, denoted by $\tilde{e} = (\tilde{e}_l, \tilde{e}_h)$, and define $\tilde{\mu}'_{chl} = \mu'(-c(\tilde{e}_h) + c(\tilde{e}_l))$ and $\tilde{\mu}'_{clh} = \mu'(-c(\tilde{e}_l) + c(\tilde{e}_h))$. Note that neither of these are 0 because profits and effort costs cannot simultaneously be equal in the two wage states, due to the symmetry of c. In the low wage state, $U'(e_l)$ just left of the discontinuity is given by

$$\lim_{e_l \to^- \tilde{e}_l} U'(e_l, w_l|\tilde{e}, \tilde{w}) = -\gamma(1 + \lambda\eta(1-p) + p\eta\tilde{\mu}'_{clh})(\tilde{e}_l - \underline{e}) - p\alpha\eta\lambda^2 b\Delta K,$$

relying on the fact that for any $\epsilon > 0$, $\pi(\tilde{e}_l - \epsilon, w_l) < \pi(\tilde{e}_l, w_l) = \pi(\tilde{e}_h, w_h)$ by hypothesis and $c(\tilde{e}_l - \epsilon) > c(\tilde{e}_l)$ by Assumption 2 and our previous observation that $\tilde{e}_l \leq \underline{e}$. Similarly, just to the right of the kink the derivative is given by

$$\lim_{e_l \to^+ \tilde{e}_l} U'(e_l, w_l|\tilde{e}, \tilde{w}) = -\gamma(1 + p\eta\tilde{\mu}'_{clh} + \eta(1-p))(\tilde{e}_l - \underline{e}) - p\alpha\eta\lambda b\Delta K.$$

In order for $U(e_l, w_l|\tilde{e}, \tilde{w})$ to have a local optimum at $\tilde{e}_l$ we thus need $p\alpha\eta\lambda^2 b\Delta K \leq -\gamma(1 + p\eta\tilde{\mu}'_{clh} + \lambda\eta(1-p))(\tilde{e}_l - \underline{e})$ and $p\alpha\eta\lambda b\Delta K \geq -\gamma(1 + p\eta\tilde{\mu}'_{clh} + \eta(1-p))(\tilde{e}_l - \underline{e})$. This requires that $\lambda(1 + p\eta\tilde{\mu}'_{clh} + \eta(1-p)) \leq (1 + p\eta\tilde{\mu}'_{clh} + \lambda\eta(1-p))$, or $\lambda - 1 \leq (1-\lambda)p\eta\tilde{\mu}'_{clh}$. But since $\lambda > 1$, the LHS is positive and the RHS is negative, so this statement is never true, and $(\tilde{e}_l, \tilde{e}_h)$ is never a PPE.

Lemma 3 *Given a stochastic wage contract (w_l, w_h, p), if the firm prefers this contract to committing to paying $\underline{w}$ with certainty, the worker's PPE is never such that $c(\tilde{e}_l) = c(\tilde{e}_h)$.*

Proof of Lemma 3

Assume towards contradiction that a personal equilibrium plan exists, denoted by $\tilde{e} = (\tilde{e}_l, \tilde{e}_h)$, such that $c(\tilde{e}_l) = c(\tilde{e}_h)$. For this to be true, $U(e_l, w_l|\tilde{e}, \tilde{w})$ must have a local optimum at $\tilde{e}_l$ and $U(e_h, w_h|\tilde{e}, \tilde{w})$ must have a local optimum at $\tilde{e}_h$. Define $\tilde{\mu}'_{\pi lh} = \mu'(\pi(\tilde{e}_l, w_l) - \pi(\tilde{e}_h, w_h))$ and similarly for $\tilde{\mu}'_{\pi hl}$. In low wage case, the value of U' just to the left of the kink at $\tilde{e}_l$ is then given by (see proof of Proposition 3 in Appendix A above):

$$\lim_{e_l \to^- \tilde{e}_l} U'(e_l, w_l|\tilde{e}, \tilde{w}) = -\gamma(1 + \eta\lambda p + \eta\lambda(1-p))(\tilde{e}_l - \underline{e}) - p\alpha\eta\lambda b\Delta K\tilde{\mu}'_{\pi lh},$$

and the value of U' just to the right of the kink is

$$\lim_{e_l \to^+ \tilde{e}_l} U'(e_l, w_l|\tilde{e}, \tilde{w}) = -\gamma(1 + \eta p + \eta(1-p))(\tilde{e}_l - \underline{e}) - p\alpha\eta\lambda b\Delta K\tilde{\mu}'_{\pi lh},$$

Putting these together, for $\tilde{e}_l$ to be a local optimum of the utility function after the low wage is realized, it is necessary that $-(1+\eta)\gamma(\tilde{e}_l - \underline{e}) \leq p\alpha\eta\lambda b\Delta K\tilde{\mu}'_{\pi lh} \leq -(1+\eta\lambda)(\tilde{e}_l - \underline{e})$. Additionally, because $c(\tilde{e}_h) = c(\tilde{e}_l) \Rightarrow \tilde{e}_l = 2\underline{e} - \tilde{e}_h$, we can rewrite this as $(1+\eta)\gamma(\tilde{e}_h - \underline{e}) \leq p\alpha\eta\lambda b\Delta K\tilde{\mu}'_{\pi lh} \leq (1+\eta\lambda)\gamma(\tilde{e}_h - \underline{e})$.

In the latter high wage case we similarly have

$$\lim_{e_h \to^- \tilde{e}_h} U'(e_h, w_h|\tilde{e}, \tilde{w}) = -\gamma(1 + \eta p + \eta(1-p))(\tilde{e}_h - \underline{e}) + (1-p)\alpha\eta b\Delta K\tilde{\mu}'_{\pi hl}$$

and

$$\lim_{e_h \to^+ \tilde{e}_h} U'(e_h, w_h|\tilde{e}, \tilde{w}) = -\gamma(1 + \eta\lambda p + \eta\lambda(1-p))(\tilde{e}_h - \underline{e}) + (1-p)\alpha\eta b\Delta K\tilde{\mu}'_{\pi hl}$$

which together mean that for $\tilde{e}_h$ to be a potential component of a personal equilibrium, it's necessary that $(1+\eta)\gamma(\tilde{e}_h - \underline{e}) \leq (1-p)\alpha\eta b\Delta K\tilde{\mu}'_{\pi hl} \leq (1+\eta\lambda)\gamma(\tilde{e}_h - \underline{e})$.

We can combine these two sets of inequalities, written as

$$(1+\eta)\gamma(\tilde{e}_h - \underline{e}) \leq \{(1-p)\alpha\eta b\Delta K\tilde{\mu}'_{\pi hl},\ p\alpha\eta\lambda b\Delta K\tilde{\mu}'_{\pi lh}\} \leq (1+\eta\lambda)\gamma(\tilde{e}_h - \underline{e}).$$

Note that the difference between the LHS and RHS is a factor of $(1+\eta\lambda)/(1+\eta)$, and the difference between the two intermediate values is a factor of $\frac{p\lambda}{1-p}\frac{\tilde{\mu}'_{\pi lh}}{\tilde{\mu}'_{\pi hl}}$. For this set of inequalities to be satisfiable, we therefore require that the latter quantity, along with its reciprocal, are both less than $(1+\eta\lambda)/(1+\eta)$.

We must break this into two cases: First, if $\pi(e_h, w_h) < \pi(e_l, w_l)$ then $\tilde{\mu}'_{\pi hl} = \lambda$ and $\tilde{\mu}'_{\pi lh} = 1$. But we can show that even if the worker would respond to a stochastic contract with a effort (e_l, e_h)

such that $c(e_h) = c(e_l)$ and $\pi(e_h, w_h) < \pi(e_l, w_l)$, the firm would never want to offer such a contract. To see this, recall that the firm's participation constraint is that $p\pi(e_h, w_h) + (1-p)\pi(e_l, w_l) > b\underline{e} - \underline{w}$. This becomes $pbe_h + (1-p)be_l - b\underline{e} > pw_h + (1-p)w_l - \underline{w}$, which is equivalent to $(2p-1)b(e_h - \underline{e}) > pw_h + (1-p)w_l - \underline{w}$ when we restrict $e_l = 2\underline{e} - e_h$. Then, note that the condition for profits to be higher in the low wage state is equivalent to $w_h - w_l > 2b(e_h - \underline{e})$ (again using $e_l = 2\underline{e} - e_h$), which we can substitute into the participation constraint to get the necessary condition that $(2p-1)b(e_h - \underline{e}) > 2pb(e_h - \underline{e}) + w_l - \underline{w} \Leftrightarrow b\underline{e} - w_l > be_h - \underline{w}$. But since $w_l \geq \underline{w}$, this can never be true.

So we are left with the possibility that the worker responds with $c(e_l) = c(e_h)$ and $\pi(e_h, w_h) > \pi(e_l, w_l)$. In this case, $\tilde{\mu}'_{\pi hl} = 1$, $\tilde{\mu}'_{\pi lh} = \lambda$, and the condition for this plan to in fact be a local optimum after the wage state is realized is that $\frac{1+\eta\lambda}{1+\eta} \geq \frac{p\lambda^2}{1-p}$ and $\frac{1+\eta\lambda}{1+\eta} \geq \frac{1-p}{p\lambda^2}$. The former is equivalent to $p \leq \frac{1+\eta\lambda}{1+\eta\lambda+(1+\eta)\lambda^2}$. Since the RHS is decreasing in λ and $\lambda > 1$, this requires that $p < \frac{1}{2}$.

But, returning to the firm's participation constraint, we have $(2p-1)b(e_h - \underline{e}) > pw_h + (1-p)w_l - \underline{w}$. Because the RHS is positive (since $w_h > w_l \geq \underline{w}$), and $e_h > \underline{e}$, this only holds if $p > 1/2$, contradicting the requirement that $p < 1/2$ for the worker to respond with these effort levels.

Proposition 5 *Given any stochastic wage contract (w_l, w_h, p) with $p \in (0,1)$, the worker's PPE $(\tilde{e}_l, \tilde{e}_h)$ must satisfy*

$$\begin{array}{ccccc} \frac{(1-p)\alpha\eta(K(w_h)-K(w_l))\mu'_{\pi hl}b}{\gamma(1+p\lambda\eta+(1-p)\eta w)} & < & \tilde{e}_h - \underline{e} & < & \frac{(1-p)\alpha\eta(K(w_h)-K(w_l))\mu'_{\pi hl}b}{\gamma(1+p\eta+(1-p)\eta x)} \\ \frac{p\alpha\eta\lambda b(K(w_h)-K(w_l))\mu'_{\pi lh}}{\gamma(1+(1-p)\lambda\eta+p\eta y)} & < & \underline{e} - \tilde{e}_l & < & \frac{p\alpha\eta\lambda b(K(w_h)-K(w_l))\mu'_{\pi lh}}{\gamma(1+(1-p)\eta+p\eta z)} \end{array}$$

where

$$(\mu'_{\pi hl}, \mu'_{\pi lh}) = \begin{cases} (1,\lambda) & \text{if } b\tilde{e}_h - w_h > b\tilde{e}_l - w_l \\ (\lambda,1) & \text{if } b\tilde{e}_h - w_h < b\tilde{e}_l - w_l \end{cases}$$

and it is never true that $b\tilde{e}_h - w_h = b\tilde{e}_l - w_l$, and

$$(w,x,y,z) = \begin{cases} (1,1,\lambda,\lambda) & \text{if } c(\tilde{e}_h) > c(\tilde{e}_l) \\ (\lambda,\lambda,1,1) & \text{if } c(\tilde{e}_h) < c(\tilde{e}_l) \\ (\lambda,1,\lambda,1) & \text{if } c(\tilde{e}_h) = c(\tilde{e}_l) \end{cases}$$

Proof of Proposition 5

We first observe that in any PPE, $\tilde{e}_h \geq \underline{e}$ and $\tilde{e}_l \leq \underline{e}$. It's trivial that $\tilde{e}_h$ cannot be strictly less than $\underline{e}$ because it would be a strict improvement, regardless of expectations, to switch to $\underline{e} + (\underline{e} - \tilde{e}_h$,

which entails the same effort cost; likewise it is trivial that $\tilde{e}_l \leq \underline{e}$.

The worker formulates a contingency plan $\tilde{e} = (\tilde{e}_h, \tilde{e}_l)$ and then finds out the true wage w_h or w_l. He then chooses effort level e_h or e_l by maximizing his utility given his plan. In the high wage state he maximizes

$$\begin{aligned} U(e_h, w_h|\tilde{e}, \tilde{w}) &= m(w_h, e_h) + EU(n(e_h, w_h|\tilde{e}, \tilde{w}) \\ &= w_h - c(e_h) \\ &\quad + p\eta\mu(-c(e_h) + c(\tilde{e}_h)) + (1-p)\eta\mu(-c(e_h) + c(\tilde{e}_l)) \\ &\quad + (1-p)\alpha\eta\mu(K(w_h) - K(w_l))\mu(be_h - w_h - b\tilde{e}_l + w_l) \\ &= -\tfrac{\gamma}{2}(1 + (1-p)\eta\mu'_{chl} + p\eta\mu'_{chh})(e_h - \underline{e})^2 + (1-p)\eta\mu'_{chl}c(\tilde{e}_l) + p\eta\mu'_{chh}c(\tilde{e_h}) \\ &\quad + (1-p)\alpha\eta\Delta K\mu'_{\pi hl}(be_h - w_h - b\tilde{e}_l + w_l) \end{aligned}$$

where

$$\mu'_{chl} = \mu'(-c(e_h) + c(\tilde{e}_l)) = \begin{cases} 1 & \text{if } c(e_h) > c(\tilde{e}_l) \\ \lambda & \text{otherwise} \end{cases}$$

and μ'_{chh} and $\mu'_{\pi hl}$ are defined similarly, and $\Delta K = K(w_h) - K(w_l)$; note that because $K(w_h) - K(w_h) = 0$, the term for reference-dependent reciprocity relative to the expected high wage state is zero.

In the low wage state, the worker maximizes

$$\begin{aligned} U(e_l, w_l|\tilde{e}, \tilde{w}) &= m(w_l, e_l) + EU(n(e_l, w_l|\tilde{e}, \tilde{w}) \\ &= w_l - c(e_l) \\ &\quad + p\eta\mu(-c(e_l) + c(\tilde{e}_h)) + (1-p)\eta\mu(-c(e_l) + c(\tilde{e}_l)) \\ &\quad + p\alpha\eta\mu(K(w_l) - K(w_h))\mu(be_l - w_l - b\tilde{e}_h + w_h) \\ &= -\tfrac{\gamma}{2}(1 + (1-p)\eta\mu'_{cll} + p\eta\mu'_{clh})(e_l - \underline{e})^2 + (1-p)\eta\mu'_{cll}c(\tilde{e}_l) + p\eta\mu'_{clh}c(\tilde{e_h}) \\ &\quad - p\alpha\eta\lambda\Delta K\mu'_{\pi lh}(be_l - w_l - b\tilde{e}_h + w_h) \end{aligned}$$

with $\mu'_{\cdot}$ defined analogously to μ'_{chl} above.

The first order conditions in the high and low wage states respectively are

$$U'(e_h w_h|\tilde{e}, \tilde{w}) = -\gamma(1 + (1-p)\eta\mu'_{chl} + p\eta\mu'_{chh})(e_h - \underline{e}) + (1-p)\alpha\eta b\Delta K\mu'_{\pi hl}$$

and

$$U'(e_l, w_l|\tilde{e}, \tilde{w}) = -\gamma(1 + (1-p)\eta\mu'_{cll} + p\eta\mu'_{clh})(e_l - \underline{e}) - p\alpha\eta\lambda b\Delta K\mu'_{\pi lh}$$

which are of course only defined where $\mu'_{\cdot}$ is defined.

Lemma 2, stated and proved above, shows that in any PPE the worker chooses $\tilde{e}_h$ and $\tilde{e}_l$ so that profits are unequal in the two wage states; therefore the optimum of the above utility functions does not occur at a kink where $\mu'_{\pi\cdot\cdot}$ doesn't exist. That is, although the firm does not earn equal

profits in the two wage states, it optimally randomizes between wage states to balance current profits against managing future wage expectations.

We therefore know that in any personal equilibrium $\mu'_{\pi hl}$ and $\mu'_{\pi lh}$ are defined and are either 1 and λ or vice versa. In the event that $c(\tilde{e_h}) \neq c(\tilde{e}_l)$, we can also assume that μ'_{chl} and μ'_{clh} are defined and are either 1 and λ or vice versa.

By definition of a PPE, the optimal choice of e_h and e_l *must* occur at the kinks where $e_h = \tilde{e}_h$ and $e_l = \tilde{e}_l$. The workers utility functions given the wage state and effort contingency plan therefore must have local maxima at these points, which means the marginal utility functions must change signs accordingly. Any PPE must therefore have the following characteristics:

$$\lim_{e_h \to {}^-\tilde{e}_h} U'(e_h, w_h|\tilde{e}, \tilde{w}) = -\gamma(1 + \eta\tilde{\mu}'_{chl}(1-p) + \eta p)(\tilde{e}_h - \underline{e}) + (1-p)\alpha\eta b\Delta K\tilde{\mu}'_{\pi hl} \geq 0$$

$$\lim_{e_h \to {}^+\tilde{e}_h} U'(e_h, w_h|\tilde{e}, \tilde{w}) = -\gamma(1 + \eta\tilde{\mu}'_{chl}(1-p) + \eta\lambda p)(\tilde{e}_h - \underline{e}) + (1-p)\alpha\eta b\Delta K\tilde{\mu}'_{\pi hl} \leq 0$$

$$\lim_{e_l \to {}^-\tilde{e}_l} U'(e_l, w_l|\tilde{e}, \tilde{w}) = -\gamma(1 + \eta\lambda(1-p) + \eta\tilde{\mu}'_{clh}p)(\tilde{e}_l - \underline{e}) - p\alpha\eta\lambda b\Delta K\tilde{\mu}'_{\pi lh} \geq 0$$

$$\lim_{e_l \to {}^+\tilde{e}_l} U'(e_l, w_l|\tilde{e}, \tilde{w}) = -\gamma(1 + \eta(1-p) + \eta\tilde{\mu}'_{clh}p)(\tilde{e}_l - \underline{e}) - p\alpha\eta\lambda b\Delta K\tilde{\mu}'_{\pi lh} \leq 0,$$

These expressions combine to yield

$$\frac{(1-p)\alpha\eta(K(w_h) - K(w_l))\mu'_{\pi hl}b}{\gamma(1 + p\lambda\eta + (1-p)\eta w)} < \tilde{e}_h - \underline{e} < \frac{(1-p)\alpha\eta(K(w_h) - K(w_l))\mu'_{\pi hl}b}{\gamma(1 + p\eta + (1-p)\eta x)}.$$

Proceeding similarly for $U'(e_l, w_l|\tilde{e}, \tilde{w})$, we find that

$$\frac{p\alpha\eta\lambda b(K(w_h) - K(w_l))\mu'_{\pi lh}}{\gamma(1 + (1-p)\lambda\eta + p\eta y)} < \underline{e} - \tilde{e}_l < \frac{p\alpha\eta\lambda b(K(w_h) - K(w_l))\mu'_{\pi lh}}{\gamma(1 + (1-p)\eta + p\eta z)}.$$

On the other hand, if $c(\tilde{e}_h) = c(\tilde{e}_l)$ the only thing that changes in the derivation above is that now $\lim_{e_h \to {}^-\tilde{e}_h} \mu'_{chh} = \lim_{e_h \to {}^-\tilde{e}_h} \mu'_{chl} = 1$ and $\lim_{e_l \to {}^-\tilde{e}_l} \mu'_{cll} = \lim_{e_l \to {}^-\tilde{e}_l} \mu'_{chl} = \lambda$ and vice versa for the upper limits. This completes all cases stated.

C Appendix: Additional analyses

	16	17	18	19	20	21	22	23	24	25
Baseline effort	1.059	1.059	1.059	1.059	1.059	1.059	1.059	1.059	1.059	1.059
	(0.157)***	(0.157)***	(0.157)***	(0.157)***	(0.157)***	(0.157)***	(0.157)***	(0.157)***	(0.157)***	(0.157)***
Full surprise	6.446	6.416	6.804	6.956	6.483	6.113	6.442	6.675	8.385	11.308
	(2.651)**	(2.629)**	(2.647)**	(2.731)**	(2.760)**	(2.903)**	(2.916)**	(2.956)**	(3.046)***	(3.202)***
Partial surprise	3.765	3.331	3.700	3.631	2.660	1.968	2.254	2.390	4.160	7.014
	(1.808)**	(1.729)*	(1.689)**	(1.800)*	(1.738)	(1.882)	(1.733)	(1.796)	(1.855)**	(2.055)***
Raise	15.836	15.836	15.836	15.836	15.836	15.836	15.836	15.836	15.836	15.836
	(5.746)***	(5.746)***	(5.746)***	(5.746)***	(5.746)***	(5.746)***	(5.746)***	(5.746)***	(5.746)***	(5.746)***
Constant	24.658	24.688	24.300	24.148	24.622	24.992	24.662	24.429	22.720	19.797
	(3.861)***	(3.919)***	(3.959)***	(4.010)***	(4.031)***	(4.103)***	(3.984)***	(3.992)***	(4.070)***	(4.377)***
R^2	0.50	0.50	0.50	0.50	0.49	0.49	0.49	0.49	0.49	0.50
N	780	780	780	780	780	780	780	780	780	780

Table 4: OLS regressions of effort levels in periods 6-25 in the Raise and SmallRaise treatments. Standard errors are robust and clustered by worker. "Full surprise" is an indicator for period 6, "Partial surprise" for periods 7 through k-1 where k is a varying cutoff indicated in the column headers, and "Raise" for the Raise treatment, while the Constant captures effort in the non-surprising periods 8-25 in the SmallRaise treatment. Controls include gender and economic student status, and controls and baseline effort are all centered to have mean zero. Statistical significance denoted at the 10% (*), 5% (**) and 1% (***) levels.