On the Power of Surprising versus Anticipated Gifts in the Workplace*

Rosario Macera†
Universidad de los Andes

Vera L. te Velde ‡
University of Queensland

January 29, 2018

Abstract

We study a largely neglected aspect of the design of gift-exchange field experiments: the surprising nature of the treatment group wage raise. We show that if reciprocal workers have expectations-based reference-dependent preferences and they expect to work at the market wage, a surprising gift can boost effort to the extent of increasing profits. The power of unanticipated gifts, however, is tighter in repeated interactions in which workers can update beliefs rationally: since workers negatively reciprocate expected but unfulfilled gifts, initially unanticipated gifts can induce long-term profit losses if they lead workers to expect further gifts probabilistically. We relate our predictions to the existing evidence and study the model's recommendations for the design of further field experiments.

JEL Classification: D030, J31, J33, M52

Keywords: Reference-dependent preferences, reciprocity, compensation, gift exchange.

*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, and Monash University for helpful comments and suggestions. Rosario Macera especially thanks Fondecyt Grant Iniciación #11140101 for funding.

†rmacera@uandes.cl, Av. Monseñor Alvaro del Portillo 12455, Las Condes, Santiago, Chile. +56 2 2618 1072.

‡v.tevelde@uq.edu.au, Colin Clark Bldg. 39 Level 6, St. Lucia, QLD, 4072, Australia. +61 7 334 67047.

1 Introduction

This paper studies the theoretical consequences of a seemingly innocuous feature of the standard gift-exchange field experiment: granting the wage increase unexpectedly. Following the pioneering work of Gneezy & List (2006), the baseline design increases the wage of a random treatment group of workers to compare their productivity to that of a control group. If reciprocal concerns drive effort provision, the wage increase triggers higher effort as workers reciprocate the gift (Akerlof 1984). To avoid selecting abler workers with higher reservation wages, however, all workers are hired at the market wage, and then the raise is granted to the treatment group right before executing the job for the first time.

This paper argues that even though it is helpful for avoiding selection, surprising workers with a wage increase is not a standard practice, and is therefore a departure from the expectations workers hold about the labor relationship. To study the consequences of this departure on effort provision and gift profitability, we develop a model of reference-dependent reciprocity in which expected outcomes constitute the workers' reference points. The idea that expectations about an upcoming outcome determine the response to the actual outcome dates back to Vroom (1964), but it was popularized by Kőszegi & Rabin (2006, 2007, 2009) who embedded expectations in a reference-dependent model with loss aversion as in Kahneman & Tversky (1979, 1991).[1] Our model follows Kőszegi & Rabin by adding loss aversion into the agent's preferences, and extends it by incorporating reference-dependent social preferences. In order to capture the beliefs induced in a gift-exchange field experiment, we also depart from Kőszegi & Rabin by considering completely unanticipated surprises, but then examine how these predictions shift when expectations are allowed to form rationally.

After presenting the model, Section 3 starts by showing that an unexpected wage raise elicits a transitorily larger effort response relative to anticipated gifts. Because a surprising wage increase is a pleasing departure from the expectation of working at the market wage, reciprocal effort can exceed what would be predicted by standard reciprocal preferences. Because expectations adapt, however, workers acclimatize to the new higher fixed wage and effort wanes back to baseline levels in the absense of further wage increases.

Gilchrist, Luca & Malhotra (2016) provide direct evidence that gifts are most effective when they are surprising. They hired three groups of oDesk workers for a one-time data-entry task, all of whom requested wages of less than \$3 per hour in their profiles. They hired the first group at \$3, the second they hired at \$4, while the third they hired at \$3 but surprised them, right before work, with a \$1 increase per hour. They found that paying \$3+\$1 yielded a 20% increase in productivity compared to paying \$4, while there were no productivity differences between the \$3 and \$4 groups. Relatedly, using a laboratory experiment, Sliwka &

[1] Vroom's 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).

Werner (2017) showed that an increasing wage profile, not communicated in advance to subjects, elicited higher effort relative to a constant wage profile with equal aggregated cost.[2]

There is also evidence that reciprocal effort in response to surprising wage raises may be transitory. In a data-entry study, Gneezy & List (2006) recruited workers for the market wage of $8 and then surprised them with a substantial permanent increase to $12. They find that the initially significant effort response of 27% falls to an insignificant 11% over a six-hour period. In a second study on door-to-door fundraising, subjects were hired at $10 per hour and then surprised with a 100% increase. Effort initially spiked by 72% but then fell back to an insignificant 6%. Similarly, Bellemare & Shearer (2009) found the effort response to a surprise bonus to workers in a tree-planting firm was concentrated on the day of the bonus only. Laboratory experiments with real-effort tasks have also replicated the waning of gift exchange. For instance, in a multiple-period experiment, Sliwka & Werner (2017) examine how reciprocal effort is affected by the timing of wage increases. They document that reciprocal effort does wane back to its baseline level in response to a surprising permanent wage increase, rendering the permanent raise ineffective at increasing overall productivity.[3]

In Section 4 we study the consequences of an initially unexpected wage raise on the effort response in a repeated principal-agent interaction in which workers update their expectations rationally, as one would expect in longer term real-world jobs. We show that, in this case, the power of surprises is not sustainable. Workers with rational expectations will now understand that the firm might grant additional gifts in a profit-maximizing manner, leading to a (stochastic) expectation of further gifts. These expectations, however, hurt firms' expected profits as they lead to retaliation in the form of lower effort when unfulfilled. Even when the firm manages expectations by granting gifts only occasionally, the retaliation always outweighs the benefits of the probabilistic gifts. Surprising gifts, therefore, are more likely to be profitable in shorter interactions in which the benefits of the surprise outweigh the longterm costs resulting from unfulfilled expectations of further wage raises caused by the workers' rational updating of expectations.[4]

[2] It could be the case, however, that subjects in Sliwka & Werner (2017) did come to expect future wage raises after the wage increase in the first period. Their result, therefore, is compatible with our predictions only if subjects did not update their gift expectations during the eight experimental periods, each lasting 250 seconds.

[3] A standard explanation for the waning pattern is fatigue: workers are just too tired to continue reciprocating after having increased effort in response to the gift. Suggestive evidence ruling out this hypothesis, however, is found in Sliwka & Werner (2017), as previously described. They find that workers paid with a sequence of small wage increases instead of a fully surprising wage increase, do not display a decreasing pattern of effort, despite the former induces higher aggregated effort at the same overall cost. Other suggestive evidence ruling out fatigue comes from comparing the effect of gifts relative to that of piece rates, where piece rates yield higher effort even late in the work period (Esteves-Sorenson 2017, DellaVigna et al. 2016). Most of the studies that do not find a waning effect are studies with short-term interactions in which subjects likely don't have time to grow accustomed to the gift. Kube, Maréchal & Puppe (2012), for example, does not observe significant waning of effort during a three-hour work period.

[4] Even though to the best of our knowledge there is no evidence in economics of monetary gifts leading to the expectation of further gifts, this is intuitive and relates to the literature of learning under ambiguity. For instance, Epstein & Schneider (2007) develop a model in which players do not know the relevant probability distributions but nonetheless update the relevant probability space. Relatedly, there is a growing literature in contract theory studying the shape of the optimal incentive scheme when workers

The result that surprising gifts can profitably increase effort levels, but that a long-term interaction limits this result, leads to several recommendations for field experiment designs. First, if the long-term effort response to gifts is of interest, the treatment group should be given time to acclimatize to the wage raise by delivering the raise news in advance in order to avoid short-term responses. Additionally, to prevent the long-term downward bias on effort provision caused by unfulfilled expectations of further gifts, field-experiment protocols should be attentive to the expectations induced by the method used to grant the gift. We discuss these and other recommendations in Section 5.

This paper makes several contributions. First, we show that the profitability of gift exchange depends on the surprising versus anticipated nature of the gift. Our model thus relates to others theoretically studying the efficiency of gift exchange. In this literature, 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, while Netzer & Schmutzler (2014) shows that gift exchange does not arise in equilibrium if agents have intentions-based reciprocal preferences as they cannot interpret profitable wage increases as kind. Unlike these papers, ours focuses on the properties of the gift itself rather than on the agents' preferences or the economic environment.[5]

Our paper also relates to a few others studying 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). They show that dictator transfers can both decrease and increase with recipients' expectations, depending on the weight put on positive and negative surprises, respectively. Unlike these papers, however, we expressly 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.

do not know the whole action space, and as a consequence, they can be surprised by actions that they were not initially aware of. See Von Thadden & Zhao (2012) and Carroll (2015).

[5] 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 agent 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 (2012*a*) and DellaVigna et al. (2016) 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.

Second, our model also contributes to another issue that has not received due attention: gift exchange in repeated interactions. After Akerlof (1982), most of the literature on gift exchange (both in the field and the few papers in the laboratory in which repeated interaction takes place) has used a permanent wage raise over a short period, which is assumed to mimic an above-market fixed wage. Even though this is necessary to avoid confounding reciprocity with reputation, this restriction has drawn attention away from the properties of gifts in repeated interactions. Our analysis enriches our knowledge of gift exchange by studying a possible alternative to a permanent wage raise, namely, random one-period gifts. We show that if firms want to motivate workers through gifts, managing expectations about future gifts is of critical importance for their success.

Third, in addition to contributing to our understanding of the profitability of gift exchange and the design of the standard gift-exchange experiment, this paper also contributes to the literature on expectations-based reference-dependent preferences by proposing a novel model of reference-dependent reciprocity. 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, the 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 (2011), and Lien, Peng & Zheng (2016) provide empirical evidence.[6] We argue that reference-dependent preferences and social preferences naturally combine to improve our understanding of gift exchange, both in field experiments and in real firms.

Lastly, our results emphasize that a crucial aspect of the architecture of an incentive scheme might lie in seemingly innocuous elements that can nonetheless affect beliefs. For instance, the way a firm communicates with workers about the change or introduction of an incentive scheme will probably influence expectations about present and future payments, in the same vein as Englmaier, Roider & Sunde (2017), who show that a simple reminder can make an incentive scheme more effective. Firms, therefore, should pay as much attention to how an incentive is implemented as to its actual payments, as implementation affects expectations, which in turn shape the response to rewards.

[6] 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.

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 assumed to be 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 or gain-loss utility from comparing actual outcomes with a reference point (Assumption 3). Define $\underline{w}$ and w_h as the market wage and an above-market clearing wage, respectively, where $w_h > \underline{w} > 0$.

Assumption 2 *(Consumption Utility $m(\cdot)$).*

(i) *Consumption utility is linear in the wage w.*

(ii) *The market wage $\underline{w}$ is the workers' reservation wage.*

(iii) *Effort costs, given by $c(e)$, are minimized at a baseline level $\underline{e} > 0$, and are convex. For simplicity, specify $c(e) = \frac{\gamma}{2}(e - \underline{e})^2$, $\gamma > 0$.*

(iv) *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) defines the market wage $\underline{w}$ as the outside option, which is used to determine the participation constraint in Section 4. Part (iii) states that the cost of effort is convex (as standard in principal-agent models) and minimized at $\underline{e}$, which captures 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 has a strictly positive interior minimum is also mathematically immaterial but made to aid with intuition. Finally, part (iv) assumes that consumption utility is additive, again, as standard in principal-agent models.

To isolate the role of surprises, we delegate the analysis including a pure reciprocity component in $m(\cdot)$ to Appendix B. 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 it introduces irrelevant complications in the analysis and results. We describe after each of the main propositions how the inclusion of baseline reciprocity affects them.[7]

Assumption 3 describes the second component of the workers' preferences: gain-loss utility relative to a reference point. In our model the two relevant domains are effort and reciprocity, the latter being the novel component. Similarly to models of intentions-based social preferences such as Rabin (1993), Levine (1998), Dufwenberg & Kirchsteiger (2004), or Falk & Fischbacher (2006), gain-loss utility from reciprocity assumes the worker puts positive weight on the employer's profits if she has been unexpectedly kind and negative weight if she has been unexpectedly unkind.

Assumption 3 *(Gain-Loss Utility $n(\cdot|\cdot)$).*

(i) *The firm's kindness $K(w)$ is strictly increasing and strictly concave in w, and it is normalized to zero at the market wage, $K(\underline{w}) = 0$.*

(ii) *The gain-loss utility function is $\eta\mu(x)$, where $\eta > 0$ is the relative weight put on gains and losses; $\mu(x)$ is piece-wise linear with a slope of 1 for $x \geq 0$ and $\lambda > 1$ for $x < 0$, where λ is the loss aversion parameter.*

(iii) *Given a reference effort level $\tilde{e}$, gain-loss utility from effort is*

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

(iv) *Given a reference wage $\tilde{w}$, a reference effort $\tilde{e}$, and a reciprocity parameter $\alpha > 0$, gain-loss utility from reciprocity is,*

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

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

Part (i) assumes that the kindness function $K(\cdot)$ is strictly increasing and concave. Concavity captures the intuition that the marginal utility of kindness is strictly decreasing but always positive, as usually assumed for positive hedonic feelings. Normalizing the kindness function to zero at the market wage reflects the evidence that workers consider the market wage as the fair wage and thus neither kind nor unkind.[8] The assumption

[7] 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).

[8] Kahneman, Knetsch & Thaler (1986) present experimental evidence of this assumption.

that the gain-loss utility function is piece-wise linear, as described in part (ii), is standard in applications of reference-dependent preferences and is made to highlight the role of loss aversion in the predictions.[9] Part (iii) and (iv) describe gain-loss utility in effort and reciprocity, the two relevant dimensions in a model with fixed wages.[10] In particular, part (iii) describes gain-loss utility in effort costs, and part (*iv*) describes gain-loss utility in reciprocity by assuming a multiplicative form for gain-loss utility in kindness and in the firm's profits.

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 who 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.

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

Assumption 4 *(Total Utility). Total utility is the sum of material and gain-loss 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 gain-loss utility, we specify how the agent builds his effort and wage expectations.

2.1) *Effort expectations.* Given his wage expectations, the agent forms his effort expectations according to the principle of the preferred personal equilibrium (Kőszegi & Rabin 2006). We restrict attention to pure strategies for the agent, which is without loss of generality until Section 4. Definition 1 formally presents this effort plan in the simplest case with only one period and deterministic $\tilde{w}$,

Definition 1 *(The Agent's Preferred Personal Equilibrium (PPE) 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*

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

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

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

[9] 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.

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

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 his plan after the formation of the plan affects his reference point. Part ii) states that, if there are multiple credible plans, then the worker chooses the one providing the highest utility. This definition follows Kőszegi & Rabin (2006)'s preferred personal equilibrium concept.[11]

2.2) *Wage expectations.* We explore the worker's effort response and the gift profitability under two different assumptions about the formation of the wage expectation, $\tilde{w}$. To replicate a standard gift-exchange field experiment, Section 3 assumes the worker is unaware of the possibility of a wage raise and thus compares this surprising gift to an expectation of being paid the wage he was hired at with certainty. Assuming the agent does not foresee a (potentially) profitable gift is the equivalent to assume 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)).[12] In Section 4 we extend the belief formation process to allow for rational updating of expectations, as might reasonably be expected after an initially surprising gift causes the agent to anticipate further potential deviations from the market wage.

(3) *Timing.* Each section specifies the details of the setting, but the general timing for a given period is as follows. The principal announces a wage. The worker then forms expectations about the wage he is going to receive, $\tilde{w}$, and forms a credible effort plan, $\tilde{e}$. Then the actual wage is revealed, and the worker exerts effort, maximizing his total utility relative to his expected wage and effort.

3 The Power of Fully Surprising Gifts

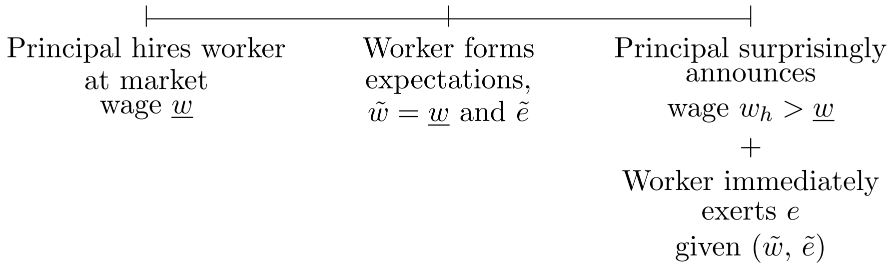
> *"If your boss walked over your desk and handed you \$10,000, would it make you work harder for the rest of the day, or the rest of the year? I think it would!"*
>
> — Member in Style.Gather.com commenting on Oprah's surprising bonus to her magazine employees in 2010.

This section replicates the framework in a one-shot gift-exchange field experiment in which subjects in the treatment group receive an entirely unexpected wage increase. To model the surprising gift, we assume that (1) the worker, being unaware that he is part of a study, believes that the actual wage will be the wage at which he was hired, and (2) because the relationship is one shot, he does not form expectations of further wage changes.

[11] We generalize Definition 1 to the stochastic reference point case in Section 4.2.

[12] 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 heterogenous 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 in a gift-exchange field experiment



Section 4 extends this analysis to the case of an ongoing relationship, in which rational workers come to expect further—possibly stochastic—gifts. Aside from the central message of gifts being more effective when they are unanticipated, we explore the size of the optimal gift, and the importance of expectations for wage cuts.

Figure 1 summarizes the timing and assumptions in the one-shot game. The principal hires the agent at the market wage, $\underline{w}$, to work for one period. Given this wage, the agent forms the belief that $\tilde{w} = \underline{w}$. Using this belief, he forms his effort expectation, $\tilde{e}$. Then the principal grants a surprising gift $w_h > \underline{w}$, the agent immediately exerts effort e, and the interaction ends.

Following the assumption that the agent forms his effort plans rationally given his wage expectations, we assume that the effort plan is credible and maximizes his ex ante utility as in Definition 1. Lemma 1 presents one simple but useful implication of the equilibrium effort whenever the reference point is deterministic,

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: since agents form effort plans rationally, any fixed wage implies the gain-loss utility is zero when expectations are met. As a consequence, 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 fully surprising gifts

As in Figure 1, suppose that the employer announces the wage gift, $w_h > \underline{w}$, immediately prior to the worker exerting effort, so that expectations do not have time to adapt. This wage news forces the agent to maximize

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

his total consumption and gain-loss utility *given* his wage and effort expectations, $\tilde{w} = \underline{w}$ and $\tilde{e} = \underline{e}$ (see Lemma 1). Namely, the worker's problem is to immediately choose an effort level e^* to maximize his utility given his (unrealized) reference point

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

To understand how this re-optimization necessarily leads to an increase in effort relative to an anticipated gift, consider the possibilities that the worker faces in this optimization problem. First, notice that he cannot simply solve the first order condition because his utility function has a kink at the effort level satisfying $\pi(e, w) = \pi(\underline{e}, \underline{w})$, where the firm's profits shift from the loss domain to the gain domain.[14] But if his optimal effort does happen to occur at a differentiable point in his utility function, then the first-order condition applies, and we can rearrange it to show that effort corresponds to

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

where the inequality holds because $\mu'_\pi = \mu'(\pi(e^*, w_h) - \pi(\underline{e}, \underline{w})) > 0$. The worker therefore exerts more effort than he would have absent the surprising gift in order to reciprocate the principal's unexpected kindness.

If the optimization process runs into the kink of the utility function, however, 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,

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

Details are relegated to the appendix in the proof of Proposition 1, but it turns out that these responses are in

[14]Notice that no such kink exists in the effort domain since *any* deviation from the plan of minimal effort implies a loss.

fact optimal so long as the profit conditions they imply hold, and when neither hold, the worker will optimally respond with $e^* = \underline{e} + (w_h - \underline{w})/b$. We have proven the following:

Proposition 1 *A worker who is expecting to receive $\underline{w}$ with certainty but who 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 an important aspect of the result: fully surprising gifts trigger more effort than had the agent anticipated the gift or had time to update his wage expectations. To see this, notice that whenever the worker anticipates the wage raise and his expectations adapt, $\tilde{w} = w_h$, he exerts $\underline{e}$ (see Lemma 1), which from Proposition 1 is strictly smaller than e^*.

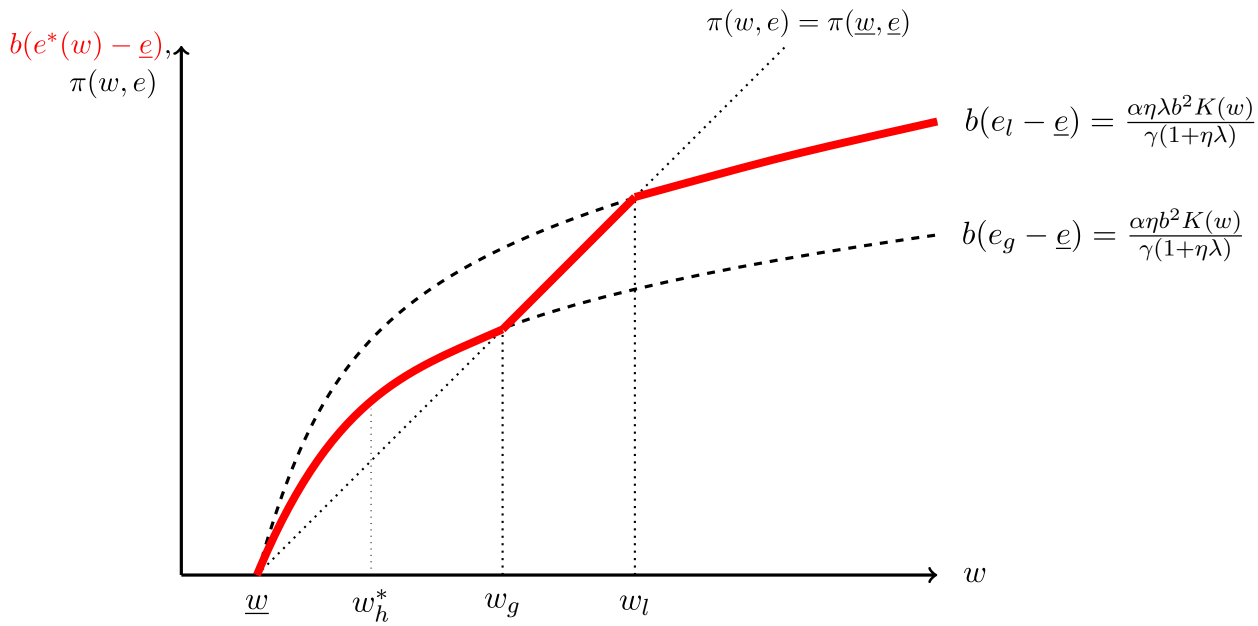
Corollary 1 *A fully surprising gift $w_h > \underline{w}$ leads to higher effort than a fully anticipated gift of the same magnitude.*

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, Appendix B presents the analysis including baseline reciprocity.

3.2 Profitable Surprising Gifts

We now turn to the profitability of a surprising gift in a one-shot interaction, as used in gift-exchange field experiments. To derive the optimal gift, Figure 2 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

Figure 2: Revenue-response function in a one-time interaction with profitable gifts.



e^*, described in Proposition 1. To 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. Intuitively, 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. Corollary 2 summarizes.

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

Once we have defined the set of profitable gifts, we can look for the profit-maximizing one. As shown in Figure 2, the optimal gift w_h^* maximizes the distance between the break-even diagonal line and the optimal revenue-response function in the range $w < w_g$. Proposition 2 formalizes this optimal gift and describes how it depends on the model parameters.

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

1. *Exists if $K'(\underline{w}) > \frac{\gamma(1+\eta\lambda)}{\alpha\eta b^2}$, and*
2. *is in the range $\underline{w} < w_h < w_g = \frac{\alpha\eta b^2 K(w)}{\gamma(1+\eta\lambda)}$ and satisfies $K'(w_h^*) = \frac{\gamma(1+\eta\lambda)}{\alpha\eta b^2}$.*
3. *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.*

Proposition 2, part 1, shows that a necessary and sufficient condition for the optimal gift to exist is that the optimal revenue-response function rises above the break-even diagonal. This occurs if the derivative of the kindness function at the market wage is large enough, i.e., 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.

Proposition 2, part 3 performs some sensitivity analysis on the model parameters. First consider the cost of effort parameter γ. 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, it decreases the size of it. 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), find that when difficult "level 5" mazes are used instead of "level-2" mazes, gift exchange is significantly reduced.[15] 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] 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.

The optimal gift responds similarly to changes in λ, which measures the impact of losses relative to gains, and γ. Because the worker does not experience losses in reciprocity in the region of profitable gift exchange, $\underline{w} < w_h < w_g$, λ only comes into play in the effort domain. 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. 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 (and without loss aversion, the effort response would continue along the lower curve). In terms of profitability, however, this part of the effort response function is irrelevant because the firm never wants to grant such large gifts. The only effect of loss aversion in equilibrium is therefore to increase the cost of reciprocation when effort costs are in the loss domain.

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 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. The empirical evidence supports 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.[16] Englmaier & Leider (2012*a*) explicitly varies the returns to effort in a gift-exchange laboratory game to show that higher returns to effort induce more reciprocation, while Englmaier & Leider (2012*b*) find a similar result in the field where they hinted high returns to effort by saying that the manager would be getting a bonus if the job was finished within the week.[17]

Finally, η behaves differently since it is a relative weight on any kind of gain-loss utility, in effort or reciprocity, compared to material (consumption) utility. However, because reciprocity is experienced exclusively as gain-loss utility while effort also affects consumption utility, increasing η increases the relative importance of reciprocity compared to effort, and therefore the gain in the reciprocity domain wins out and causes w_h^* to rise.

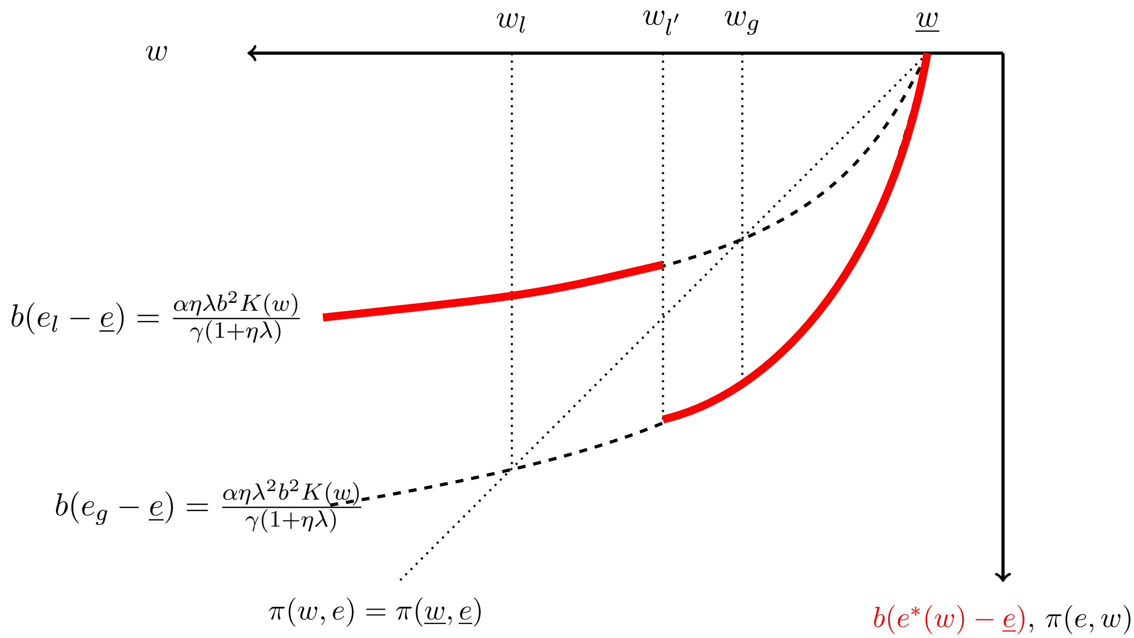
3.3 The effort response to a fully surprising wage cut

In this section we show that awareness about a wage cut matters: managing expectations prior to a wage cut can ameliorate the effort decrease induced by negative reference-dependent reciprocity. This is important as,

[16]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.

[17]DellaVigna et al. (2016), in the largest gift-exchange field experiment to date, also systematically vary the returns to the employer. They find that workers do exert more effort when the return for the employer in the stuffing-envelope task is positive, although their response is not sensitive to the precise size of the return. However, since they cannot document a significant effort response to a monetary gift, they are unable to test whether the return for the employer affects gift exchange in the field.

Figure 3: Revenue-response function to a fully-surprising wage cut in a one-time interaction.



unsurprisingly, when firms cut wages, workers frequently respond by reducing effort (Lee & Rupp 2007, Krueger & Mas 2004, Kube, Maréchal & Puppe 2013).[18]

To formalize the role of full surprises in the response to a wage cut, we extend K to negative values. In accordance with prospect theory, which holds that people have diminishing sensitivity to both gains and losses, we assume that K is convex over $\mathbb{R}^-$. In particular, for simplicity, assume that K is rotationally symmetric around $\underline{w}$: $K(\underline{w} + w) = -K(\underline{w} - w)$ for all w.

The analysis proceeds analogously to Section 3.1. The worker anticipates $\underline{w}$ and plans to exert effort $\underline{e}$, but he is surprised by a wage cut $w_c < \underline{w}$. With this new information, the worker's problem is to immediately choose an effort level e^* to maximize his utility given his (unrealized) reference point. This e^* solves

$$e^* \in \underset{e}{\operatorname{argmax}}\, w_c - \frac{\gamma}{2}(e - \underline{e})^2 + \eta\mu(-\frac{\gamma}{2}(e - \underline{e})) + \alpha\eta\mu(K(w_c) - K(\underline{w}))\mu(b(e - \underline{e}) - (w_c - \underline{w})) \tag{8}$$

where $K(w_c) - K(\underline{w}) < 0$. Depending on whether the worker reduces effort enough to hurt the firm's profits on net (e_l) or not (e_g), the two potential interior solutions to this optimization problem (not at the kink) are derived from the first-order condition of this utility function. These effort levels, and the wage cut sizes for

[18]The effort reduction in response to a wage cut does not seem to depend on the nature of the employer-employee relationship, however. Chen & Horton (2016) show that wage cuts harm effort even in online labor markets where the employment contract resembles a spot contract more than a labor contract.

which they indeed correspond to local maxima, are,

$$e_g - \underline{e} = \frac{\alpha\eta\lambda K(w_c)b}{\gamma(1+\eta\lambda)} \qquad \Leftrightarrow \qquad \frac{\alpha\eta\lambda K(w_c)b^2}{\gamma(1+\eta\lambda)} > w_c - \underline{w} \tag{9}$$

and

$$e_l - \underline{e} = \frac{\alpha\eta\lambda^2 bK(w_c)}{\gamma(1+\eta\lambda)} \qquad \Leftrightarrow \qquad \frac{\alpha\eta\lambda^2 K(w_c)b^2}{\gamma(1+\eta\lambda)} < w_c - \underline{w}. \tag{10}$$

Figure 3 shows with a thick red line the worker's response in terms of decreased revenue $b(e - \underline{e})$ as a function of the size of the wage cut. As before, the diagonal line is the break-even profit line where the decrease in revenue equals the savings from the cut; above the line it is profitable to cut wages and below the line it is not. The first dotted concave curve (closer to the x-axis) shows the revenue decrease that occurs when the worker responds with e_g, as defined in equation (9). Since he only does this when profits increase as a result of the wage cut, he follows this curve as long as it is below the diagonal, or equivalently, if the wage is smaller than the threshold $w_g = \underline{w} + \frac{\alpha\eta K(w_g)b}{\gamma(1+\eta\lambda)}$, which makes the inequality in (9) hold with equality. Similarly, the second dotted concave curve (the one further below from the x-axis) shows the revenue decrease that occurs when the worker chooses e_l, as defined in equation 10. Since he only does this when this revenue does not compensate for the cost of the gift, he follows this curve as long as he is in the loss domain, i.e., below the diagonal line or equivalently, if gifts are greater than a threshold $w_l = \underline{w} + \frac{\alpha\eta\lambda K(w_l)b}{\gamma(1+\eta\lambda)}$, which makes the inequality in (10) hold with equality. Since $w_l < w_g$, notice that—to the contrary of Figure 2—for wages between w_l and w_g, both e_l and e_g, are local best responses. The proof of Proposition 3 shows that for this range there exists a $w_{l'}$, $w_l < w_{l'} < w_g$, such that e_g is optimal when $w_c < w_{l'}$ and e_l is optimal when $w_c > w_{l'}$. Thus, the agent never prefers effort at the kink $e = \underline{e} - (\underline{w} - w_l)/b$ and the decreased revenue is discontinuous at $w_{l'}$.

The discontinuity in this response function is the key qualitative difference between the wage raise and the wage cut cases. In the wage raise case, small gifts are easily reciprocated, but as the wage rises and reciprocation becomes more expensive, workers are strongly dissuaded from switching to the loss domain by the loss aversion parameter, which increases the weight on reciprocation when in the loss domain. Workers therefore toe the line by exactly reciprocating profits until the wage rises enough that even reciprocating that much is too expensive to be worthwhile. On the other hand, in the wage cut case, small cuts are easily punished. When the cut gets large enough, the worker no longer wants to fully punish the cut, and switching to partial punishment *reduces* the weight on reciprocity, creating the discrete drop in punishment at that point.

The resulting revenue response function in Figure 3 shows that, whatever the size of the wage cut, the worker responds with a drop in effort. But had the wage cut been anticipated, Lemma 1 would have applied and effort would not have decreased at all. Proposition 3 summarizes:

Proposition 3 *A worker who is expecting to receive $\underline{w}$ with certainty but who surprisingly receives $w_c < \underline{w}$ and does not have time to update his expectations, responds with effort*

$$e^* = \underline{e} - \begin{cases} \frac{\alpha\eta\lambda^2 bK(w_c)}{\gamma(1+\eta\lambda)} & \text{if } w_c > w_{l'} \\ \frac{\alpha\eta\lambda K(w_c)b}{\gamma(1+\eta\lambda)} & \text{if } w_c < w_{l'} \end{cases}$$

where $w_{l'} = \underline{w} + \frac{\lambda(\lambda+1)}{2}\frac{\alpha\eta b^2 K(w_l)}{\gamma(1+\eta\lambda)}$ and w_l solves $w - \underline{w} = \frac{\alpha\eta\lambda K(w)b}{\gamma(1+\eta\lambda)}$ for w.

As before, in order to focus on the effect of surprises, we have not included non-reference-dependent reciprocity in levels. Nonetheless, Proposition 3 is qualitatively robust to the inclusion of such preferences; negative reciprocity is made worse by the surprise factor. Indeed, if a surprising wage cut occurs when workers are being paid above-market wages and exerting above-minimum effort, retaliation is even cheaper and more damaging for the firm. These issues are discussed in Appendix B for interested readers.

We can also compare the size of the effort response between raises and equivalently sized cuts. This is most easily done by comparing Figure 2 to Figure 3 (rotated 180°). Note that the point labeled w_l in Figure 2 exactly corresponds to the point labeled w_g in Figure 3. Then it is easy to see that small wage raises provoke a smaller effort response than equivalently sized cuts, but large wage cuts and raises are responded to in equal magnitude.[19] Corollary 3 summarizes the effect of surprising wage cuts relative to anticipated cuts (analogously to Corollary 1) and relative to surprising raises.

Corollary 3 *Retaliation against wage cuts is worse for surprising than for anticipated cuts. Furthermore, surprising wage cuts are reciprocated at least as strongly as equivalently sized surprising wage increases.*

The predicted asymmetry between wage cuts and wage increases in corollary 3 is well grounded in the experimental evidence. Hannan (2005) modifies a standard laboratory gift-exchange experiment to add an exogenous shock to the firm's profit after which firms and workers can change their wage and effort choices. She finds that adjusting wages downwards has a negative impact on effort choice, which is twice as large as the effect of a wage increase of the same magnitude. Field evidence also supports this asymmetry. Kube, Maréchal & Puppe (2013) hired workers for a data-entry task for a "projected" wage of 15 Euros. Right before work, one

[19] The asymmetric response of effort to wage increases vis-a-vis wage decreases is well established in the empirical literature (Offerman 2002, Al-Ubaydli & Lee 2009, Kube, Maréchal & Puppe 2013). Moreover, this asymmetry has also been established in prices in general (Ahrens, Pirschel & Snower 2017).

group of workers received a wage cut to 10 Euros and the other group a wage raise to 20 Euros. They found that cutting the payment reduced average output by 20% relative to the control that received the expected 15 Euros, while the high wage group did not exhibit increased effort even though effort did respond positively to monetary incentives. [20] These results are in line with the well established stylized fact that firms are reluctant to cut wages to avoid hurting workers' "morale" (e.g., Bewley (2009)).

4 Surprising Gifts in Long-Term Relationships

In this section we explore the consequences of gift anticipation. Even if workers can initially be completely surprised by a gift, as analyzed in the previous section, long-term employee-employer relationships intuitively allow workers enough time to learn how to rationally form expectations regarding the likelihood of additional gifts. Analyzing this long-term interaction therefore accomplishes two important goals, in addition to conforming to traditional rational expectations-based analysis of labor markets: First, this analysis provides the results of interest to any real-world firm with regular (non-temporary) employees. To this end we show that the benefits of gifts extend only to the short-run, and in the long-run can backfire as workers negatively reciprocate any expected further gifts. Second, the analysis tells us about the external validity of the standard gift exchange field experiment. As Section 3 showed that a surprising wage raise is not an innocuous design choice in such experiments, this section shows that short-run experiments may systematically overestimate the profitability of gift exchange. These implications of the model are discussed in more detail in Section 5.

Section 4.1 analyzes an initial benchmark case in which, despite repeated interaction, the firm is able to grant a permanent wage raise while still managing expectations about future raises effectively. In this setting, which is relevant to the interpretation of some existing field evidence on gift exchange, workers merely adapt to the new wage without forming expectations of additional raises.

Section 4.2 then conforms to standard game-theoretic analysis of belief formation in labor markets by enforcing rational expectations. This conceptually captures the long-run impact of gift exchange, after workers have learned the firm's stochastic gift granting strategy. This reveals the curse of gifts: even if a firm grants gifts only occasionally, with an optimally chosen probability, retaliation against unfulfilled expectations outweighs the benefits of positive reciprocity.

Section 4.3 explicitly models the tradeoff inherent in granting an initially surprising gift: although this first gift can profitably evoke positive reciprocity, the gift also reveals the firm's inability to commit to a particular

[20] To test whether the lack of positive reciprocity was due to workers being unable to reciprocate, they hired workers for a piece rate and verified that there was room for a productivity increase above the baseline. Workers, however, were recruited under the piece rate, opening the possibility that the subjects working for the piece rate where more productive workers relative to those hired for a fixed wage.

wage. Workers therefore infer the firm's optimal gift-giving strategy and come to expect further gifts with some profitability, as in Section 4.2. Despite the pessimistic message of Section 4.2, firms may still be able to make profitable use of gifts in sufficiently short (finite, or indefinite but with a sufficiently small discount factor) employment relationships even if they are unable to manage expectations effectively.

4.1 The Power of Gifts Can Wane 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.

In this section, we begin our analysis of long-term gift exchange by considering the simple situation in which a firm is able to grant a surprising permanent wage increase without causing the agents to expect further raises. These expectations would be correct, for instance, if 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.

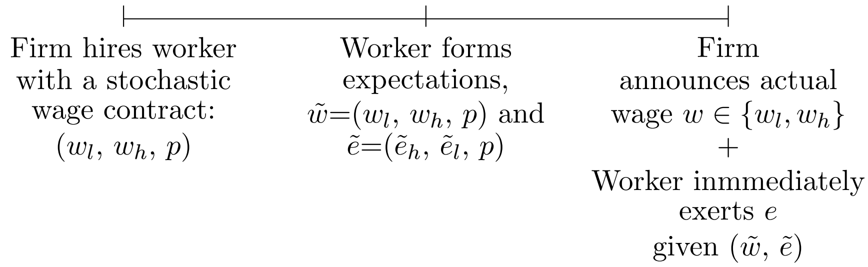
There are infinite periods. In period zero the principal hires the agent at the market wage $\underline{w}$ for all upcoming periods. Given this wage, the agent forms the belief that he will be paid the market wage for all periods, $\tilde{w}_t = \underline{w}$. By Lemma 1 the agent plans to exert minimum 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 $t > 1$.[21] 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 further surprises, then, effort returns to baseline for all future periods. Corollary 4 summarizes:

Corollary 4 *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 $t > 1$.*

As discussed in more detail in Appendix B, 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

[21] 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 stimuli 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.

Figure 4: Timing and assumptions of the stage game with an stochastic gift.



to a higher level for $t > 1$ as well. Notice that the time required for 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.[22]

4.2 Gifts are cursed: Grant them forever or never

"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.

We now explore effort provision in a repeated employment interaction in which agents are able to accurately learn their employer's gift-giving strategy and to form rational expectations accordingly. Because our analysis of surprising gifts indicates potential for profitable reciprocity, firms may infer that providing gifts repeatedly could be a profitable strategy that takes advantage of the benefits of surprising agents—at least probabilistically—to elicit higher effort. We show, however, that this intuition is incorrect: gifts are cursed whenever they are anticipated, even probabilistically, because the decrease in expected revenue due to retaliation when workers do not receive the gift outweighs the expected revenue boost that occurs when the gift is granted.

The analysis in this section focuses on the stationary equilibria of the infinite principal-agent interaction. In a stationary equilibrium the principal offers the same (potentially stochastic) compensation scheme in every

[22] In a laboratory experiment varying the time gap that subjects are given between payment information arrives and the execution of a task measuring loss aversion, Song (2016) finds that loss aversion is decreasing in the time gap. Thakral & Tô (2017) provide 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.

period. Considering non-stationary contracts—e.g., by allowing each period 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 that no dynamic considerations are needed to motivate workers.

Figure 4 summarizes the stage game timing and assumptions. The stage contract offered by the firm is now stochastic: she will pay a gift wage w_h with probability $p \in (0, 1)$ — the optimal fixed wage ($p = 0$ or $p = 1$, equivalently) when agents have rational expectations is discussed in Section 4.3. The baseline non-gift wage is w_l, paid with probability $1 - p$, where $w_h > w_l \geq \underline{w}$. The firm chooses p, w_l, and w_h optimally in order to maximize profits contingent on the agent's induced expectations.[23,24] 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 contingency 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.

Since the gift is now possibly stochastic, Definition 2 formally extends the equilibrium plan and effort decision in Definition 1 to consider stochastic reference points and contingent effort plans. In Definition 2, planned effort contingent on the wage expectations must maximize utility given the wage realization, and the full contingency plan $(\tilde{e}_l, \tilde{e_h})$ must maximize expected utility among all such credible plans. Further, as in Kőszegi & Rabin (2006), we assume that realized outcomes are compared to all possible unrealized outcomes and the overall sense of gain or loss is the probability-weighted average of these comparisons.

Definition 2 *(The Agent's Preferred Personal Equilibrium (PPE) with stochastic wage expectations)*
Given an stochastic wage expectation $\tilde{w}$=(w_l, w_h, p), a contingent effort plan $(\tilde{e}_l, \tilde{e_h})$ constitutes a preferred personal equilibrium (PPE) iff

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

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

[23] 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 but 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.

[24] 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.

(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)$\}$.

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.[25]

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 she would get by offering the market wage with certainty. Strictly speaking, this baseline for comparison is somewhat arbitrary, but the analysis is in anticipation of the following subsection in which we conceptualize a firm that explicitly chooses between preserving the illusion of an ability to commit to the market wage, by never deviating from this fixed wage, and giving up the ability to manage expectations in exchange for the short-run profits discovered in Section 3. By narrowing down in Lemmas 2 through 4 what the stochastic contract must look like, we will show that this is in fact never possible.

First, in Lemma 2 we formally state the straightforward result that workers do in fact positively reciprocate in the high-wage state and negatively reciprocate in the low-wage state. The PPE concept requires that the worker plan to reciprocate, and by the definition of PPE he then executes this plan when w_l or w_h is realized.

Lemma 2 *Given any stochastic wage contract (w_l, w_h, p), the worker's PPE $(\tilde{e}_l, \tilde{e}_h)$ satisfies $\tilde{e}_l < \underline{e}$ and $\tilde{e}_h > \underline{e}$.*

Next, we find that any stochastic contract of this form will be responded to by workers in a way that ensures unequal profits in the two wage states; that is, the worker prefers to stay away from the kink in his utility function where profits are equal in all states and thus equal to expectations in all states. The firm will therefore always strictly prefer either the high or the low wage, and thus will not be willing to randomize between them in an isolated interaction. The firm is not forced, via commitment, to randomize, but optimally randomizes despite unequal current profits in order to manage workers' expectations about the likelihood of future gifts.[26] Intuitively, while we are looking for stationary equilibria within the stage game, the repeated nature of the interaction is what incentivizes the firm to choose and stick to p despite short-run incentives to defect from this strategy.

[25]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 utility.

[26]This dependence of payments on beliefs places this model in the realm of psychological game theory (Battigalli & Dufwenberg 2009, Geanakoplos, Pearce & Stacchetti 1989).

Lemma 3 *Given a stochastic wage contract* (w_l, w_h, p)*, the worker's PPE is never such that* $\pi(e_l, w_l) = \pi(e_h, w_h)$*.*

Next we show 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.

Lemma 4 *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* $c(\tilde{e}_l) = c(\tilde{e}_h)$*.*

Having shown any desirable equilibrium will not occur at the kinks in either the effort or the profit domains, we can finally establish the main result,

Proposition 4 *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, she is left with no option but to commit herself 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 wage once she has revealed her 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 5.

Intuitively, loss aversion plays an important role in understanding Proposition 4. In earlier results, the loss aversion parameter λ folds into the overall cost of reciprocation by effectively amplifying the effort cost function, but does not otherwise qualitatively affect the firm (except if λ is large enough to preclude the existence of profitable gifts). In this repeated interaction, however, it prevents the firm from exploiting the power of gifts repeatedly. Workers expect a gift with some probability, so a realized gift puts the firm's kindness solidly in the gain domain, and an unrealized gift puts it solidly 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 4. 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.

4.3 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 is able to surprise the agent with a gift, it might be profitable even if it generates the expectation of further gifts. To this end, we assume that in the first period the agent is surprised by a profitable wage increase and that as a consequence, in period two onwards, he knows that the employer is unable to commit to a wage and rationally infers the probability of a further gift. 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.[27]

The timing and assumptions combine the unanticipated gifts of Section 3 with the long run interaction of Section 4.2 in order to mimic real belief formation in labor markets as closely as possible. 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 for all upcoming periods, $\tilde{e}_t = \underline{e}$. 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. This initial gift reveals the firm's inability to commit to a wage, and so 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 all of the following periods the principal then announces the actual wage from the set $\{\underline{w}, w_h\}$ and then the worker exerts effort, following Definition 2.[28]

If the game is infinite, Proposition 4 shows that once the firm has revealed 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 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, so 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

[27]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 dollar 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*). As explained in footnote 12, our preferred interpretation of surprises with rational agents is an asymmetry of information about the principal's action set. Intuitively, workers at *O* magazine were fully surprised by the wage increase, not because they failed to see that such is raise is profitable for the firm, but simply because they were not aware that such a gift was feasible. This is intuitive given that in the labor market these gifts are uncommon (so uncommon that that they make it to the news!)

[28]Notice we do not explicitly assume that the agent updates expectations at the end of periods $t > 1$. As explained in Section 4.2, 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 is able to profitably grant gifts under this restriction, she will certainly be able to with flexible w_l.

from an initial surprising gift therefore depends on the comparison between the benefits triggered by the surprise (Proposition 2) against 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 sufficiently low, or equivalently, if there is a sufficiently high probability each period that the work relationship will be discontinued. This proves the following corollary:

Corollary 5 *In an infinite game, gift exchange is only profitable if the firm has a sufficiently high discount factor (or equivalently, if there is a sufficiently high probability of the game ending in each period.)*

In this infinite interaction, however, a value of $p \in (0,1)$ is sustainable only because the firm must manage expectations about future gifts. The firm is unable to commit to a particular probability of a gift, but optimally chooses this probability so as to balance the gains from immediate gift exchange against the losses associated with workers learning to expect a higher probability of the gift, provided that a full surprise was feasible. Lemma 3 shows the firm always strictly prefers one of the two wage levels in the current period, so the only reason it is willing to randomize between them is that maintaining a particular probability optimally balances these factors in the long run.

In finite interactions, this expectations management is no longer relevant and reasoning unravels through backwards induction. In the final period, the equilibrium belief p must be zero or one, since we know from Lemma 3 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 at all. 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 as well. Proposition 5 shows that the optimal wage level to which the firm can credibly commit is in fact the same as the optimal fully surprising gift found in Proposition 2.

Proposition 5 *In a finite game, if the firm chooses to grant an initial profitable surprising gift, she grants the optimal gift w_h^* described in Proposition 2 in the first round, and then turns the gift into a permanent raise in later periods.*

Similarly to Corollary 5, Proposition 5 also 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 having to pay 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 is.

Again, our model abstracts from non-reference-dependent reciprocity in order to focus on the effect of surprises. Clearly if workers are reciprocal at baseline, a positive correlation between wages and effort may exist in the long run, which could improve the profitability of gift exchange even when gifts are anticipated. When gifts are stochastic, however, baseline reciprocity can render gifts even less profitable relative to credibly committing to a fixed wage. If the worker is reciprocating with above-minimal effort to begin with at some wage $w_l > \underline{w}$ (perhaps chosen by the firm to exploit non-reference-dependent reciprocity), a surprising gift $w_h > w_l$ is now more difficult to reciprocate because the worker is starting off at a higher marginal cost of effort. Conversely, unfulfilled expected gifts are extremely cheap to negatively reciprocate, as the worker can simply withdraw his previous high effort. The risk to the firm from granting surprising gifts is thus amplified when baseline reciprocity is assumed. See Appendix B for details.

5 Discussion

Before concluding, we discuss lessons and implications of our framework for the design of field experiments on gift exchange and the use of gifts to motivate workers in the labor market. Whenever possible, we present venues for future research.

5.1 Be careful with expectations to make the most of 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.

By extending the approach in Kőszegi & Rabin (2006) to model workers with reference-dependent reciprocal preferences, our model predicts that the most powerful gifts are those that are surprising, but surprises still only trigger temporary excess effort. Moreover, if surprises create the expectation of further potential gifts, they will harm the firm's future expected profits if unfulfilled. Initially profitable gifts, therefore, are cursed in repeated interactions.

These theoretical predictions suggest that field experiments on gift exchange should explicitly consider the effects that the design has on worker's wage expectations. First, to preserve the external validity of the experimental setting, subjects could be given time for their wage expectations to adapt. Gift news, for example, could be communicated to treated subjects in advance, or the effort response could be assessed in longer time spans if contamination is a concern. Lengthening the study period has the additional benefit of providing more evidence for the duration of gift exchange.[29] Section 4.2, however, indicates that more extended periods may amplify the risk that workers will start anticipating further gifts. Retaliation against unfulfilled expectations could be misinterpreted as a failure to reciprocate the initial gift, leading to an underestimated average reciprocal response. Experimental designs could mitigate this negative effect by emphasizing the one-time nature of the raise in the wording used to grant the gift.

These recommendations for leveraging expectations in field experiments open interesting questions about the mechanisms that real firms use to influence workers' expectations. Even though various aspects of the implementation of the gift may be relevant for how workers form their payment expectations, the economics literature has largely neglected these issues. For instance, one important dimension is how to communicate the gift to workers. This was clear to Disney CEO Bob Iger when he announced a $1,000 cash bonus that was explicitly and repeatedly emphasized to be one-time. From the opening quote above, however, avoiding expectations about further gifts was something Starbucks forgot to think when they granted an unusual Christmas bonus to their employees.[30]

5.2 Be thoughtful about the size of the gift.

In addition to finding that gifts are more powerful when surprising, our model also shows that only gifts that are small enough can be expected to be profitable. This is clearly relevant to experiment designs that attempt to study profitability of gift exchange. Even though how small is small naturally depends on the details of the task and the job, in current field studies there is no usually no justification for the chosen gift size. As a consequence, even with similar designs (similar population and tasks), very different gift sizes have been used. For instance, termGneezy & List (2006) granted a 100% raise to data-entry workers, while Englmaier & Leider (2012*b*) and Kube, Maréchal & Puppe (2013) granted 39% and 20% raises, respectively. See DellaVigna et al. (2016) for a detailed summary of the gift sizes used in field experiments.

The underlying reason why only small gifts can be profitable also points to other factors that may be usefully communicated indirectly through the size of the gift. Proposition 1 shows that large gifts are too costly to be fully reciprocate and thus hurt the principal's profits. Absent explicit information about profits, as it is the case in field studies, a small gift can act a signal that the gift is profitable and thus encourage reciprocity. This is consistent with existing laboratory evidence suggesting that information about the profitability of the gift matters for reciprocal behavior. Hennig-Schmidt, Rockenbach & Sadrieh (2010) show that explicit cost and

[29]Even though existing evidence (Gneezy & List 2006, Bellemare & Shearer 2009) is consistent with the waning pattern of effort predicted in Section 4.1, most experiments have a duration too short to discuss long-run effects of gifts.

[30]Read Disney's bonus tory here. Notice that another major part of the implementation of monetary gifts has to do with taxes: surprising bonuses are considered discretionary payments and thus excluded from overtime pay calculation.

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 worker's 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. Finally, Englmaier & Leider (2012*a*) find that gift exchange increases with the the payoff the firm derives from the worker's effort, although this effect is mediated by the worker's ability.[31]

5.3 Consider gift giving not only in the monetary realm.

One pending question in the gift-exchange literature is why gifts in kind trigger reciprocal effort (Maréchal & Thöni 2016) and even to a greater extent than monetary gifts, even though they are less valued (Kube, Maréchal & Puppe 2012).[32] In the context of our model, one possible answer is that, in contrast to monetary gifts, gifts in kind do not create the expectation of a further gift as their usage is usually more sporadic. This would make sense if gifts in kind are inherently less repeatable (e.g. giving employees a company sweater is something that can only reasonably be done every couple of years) or if they are tied to specific, credibly nonreplicable, events, such as hiring and promotions or holidays.

Kube, Maréchal & Puppe (2012) 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. An explanation along these lines is certainly required to understand the stronger short-run impact of gifts in kind. But our long-run analysis raises the question of whether gifts in kind might be more attractive to firms in addition for their role in managing expectations. Our theory thus predicts that this property of gifts in kind might be at the heart of the widespread use of this type of incentives in the workplace and posits a call for further research on the role of expectations for the optimality of non-monetary gifts.

5.4 Invest in making the labor relationship credible.

The major theme of this discussion is that all aspects of a gift and its implementation should be considered with regard to their impact on workers' expectations. In the design of field experiments, in particular, expectations are mediated by the credibility of the labor market. Our propositions show that an unconditional gift given

[31] The field evidence on this domain is less clear. Englmaier & Leider (2012*b*) use a data-entry task to show that a wage raise from \$13 to \$18 increases productivity only when subjects know that those who hired them will receive a bonus if the job meets a deadline. DellaVigna et al. (2016), using a much larger sample, however, do not find that reciprocal effort is responsive to the employer returns in a folding-envelopes task where the employer is a charity whose returns are varied by a matching donor.

[32] In a large-scale field experiment, however, DellaVigna et al. (2016) do not find workers reciprocate gifts in kind, even though they do not find they reciprocate monetary gifts either.

before any work may be interpreted by workers as an out-of-equilibrium play by the firm, making it difficult for workers to form reasonable expectations.

The surprising nature of the gift, however, is not the only design factor that may affect credibility in this way. Gift sizes that are incompatible with profit maximization may also distort expectations, as discussed above. The task used, the recruitment strategy, selection criteria, and the work environment are other possibilities. Further research is needed to identify how labor relationships that might fail to be credible hurt reciprocal effort, but some results indicate they are in fact impactful: Macera & te Velde (2017) show that effort exertion is dependent on whether the firm is explicit in the wording used to grant the gift that effort is expected in return. One proposed channel for such an effect is that workers find it difficult to form second-order beliefs in gift-exchange field experiments due the firm's out of equilibrium play. Even if workers form expectations, it is difficult for the experimenter to anticipate how this happens. For example, a surprising gift may convey that the employer is trying to compensate for an unexpectedly unpleasant job, 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. Observed negative reciprocity cannot be interpreted as such in the context of these unobservable expectations, and so identification requires carefully credible implementation of the labor market.[33]

We conclude that one of the future challenge of field tests in gift exchange is to employ designs that consider the equilibrium play of the firm to achieve identification, which our model indicates is required for external validity. This implies that future experimental designs should not only be guided by models that focus on the worker's response to gift giving, but which also focus on the equilibrium behavior of firms, in order to ensure the credibility of gift-giving in the field.

[33]Lack of credibility of the gift and labor relationship due to the out-of-equilibrium play of the firm could also 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 (2017) and DellaVigna et al. (2016) 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 (2012*b*) 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 (2017) find no significant gift exchange with a data-entry task despite larger than usual gifts, while DellaVigna et al. (2016) also fails to find reciprocal effort in response to monetary gifts in a stuffing-envelope field experiment with an impressive sample size. This disparity has lead to a heated debate. Levitt & List (2007) catalogues 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.

6 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 suggests that the fully surprising nature of the gift can trigger a transitorily higher effort response than that of workers in real-world firms where above-market wages are common. We show, however, that this boost in effort might not extend to a repeated interaction: deviations from the workers' rational expectation of working at the market wage can lead to future profit losses if the deviation from the initial agreement leads workers to expect additional wage raises. These results highlight that future field tests of gift exchange should be careful about unintended effects of the treatment (a surprising wage raise) on beliefs, which can affect the estimation of gift exchange profitability.

Based on this predictions, we derive several recommendations for future field tests on gift exchange concerning the timing of the gift and the wording used to deliver the gift. These recommendations and other detailed in Section 5 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, which we do not touch upon 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 view our model and those of gift exchange in general as a complementary explanation to efficiency wages rather than a substitute to classical mechanisms behind the optimality of above-market wages. Shapiro & Stiglitz (1984) proposed that it can be optimal for firms to pay above-market clearing wages to increase workers' opportunity cost of losing the job; Stiglitz (1974) and Salop (1979) argued that paying above market wages is profitable as it saves firms the onerous costs of labor turnover, while Weiss (1980) proposed that paying higher wages leads to selection of better workers as ability and reservation wages are correlated. As gift exchange builds on the agents' preference structure, it can, along the classical explanations, provide a more complete picture of the mechanisms explaining why workers' marginal productivity might depend on the wage (Yellen 1984).

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.

Ahrens, Steffen, Inske Pirschel, and Dennis J. Snower. 2017. "A theory of price adjustment under loss aversion." *Journal of Economic Behavior and Organization*, 134: 78–95.

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, and Min Sok Lee. 2009. "An experimental study of asymmetric reciprocity." *Journal of Economic Behavior and Organization*, 72(2): 738–749.

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.

Battigalli, Pierpaolo, and Martin Dufwenberg. 2009. "Dynamic psychological games." *Journal of Economic Theory*, 144(1): 1–35.

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.

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

Bewley, Truman F. 2009. *Why Wages Don't Fall During a Recession.* Harvard University Press.

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.

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).

Carroll, Gabriel D. 2015. "Robustness and linear contracts." *American Economic Review*, 105(2): 536–563.

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.

Chen, Daniel L., and John J. Horton. 2016. "Are online labor markets spot markets for tasks? A field experiment on the behavioral response to wage cuts." *Information Systems Research*, 27(2): 403–423.

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.

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. 2016. "Estimating social preferences and gift exchange at work." Working Paper.

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, Florian, Andreas Roider, and Uwe Sunde. 2017. "The Role of Communication of Performance Schemes: Evidence from a Field Experiment." *Management Science*, 63(12): 4061 – 4080.

Englmaier, Florian, and Stephen G. Leider. 2012*a*. "Gift exchange in the lab: It is not (only) how much you give..." Working Paper.

Englmaier, Florian, and Stephen G. Leider. 2012*b*. "Managerial payoff and gift exchange in the field." CESifo working paper: Behavioral Economics No. 3707.

Epstein, Larry G., and Martin Schneider. 2007. "Learning under ambiguity." *Review of Economic Studies*, 74(4): 1275–1303.

Esteves-Sorenson, Constança. 2017. "Gift exchange in the workplace: Addressing the conflicting evidence with a careful test." Working Paper.

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.

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.

Geanakoplos, John, David Pearce, and Ennio Stacchetti. 1989. "Psychological games and sequential rationality." *Games and Economic Behavior*, 1(1): 60–79.

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. 2005. "The combined effect of wages and firm profit on employee effort." *Accounting Review*, 80(1): 167–188.

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.

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.

Krueger, Alan B., and Alexandre Mas. 2004. "Strikes, scabs, and tread separations: Labor strife and the production of defective Bridgestone/Firestone tires." *Journal of Political Economy*, 112(2): 253–289.

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.

Lee, Darin, and Nicholas G. Rupp. 2007. "Retracting a gift: How does employee effort respond to wage reductions?" *Journal of Labor Economics*, 25(4): 725–761.

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.

Maréchal, Michel André, and Christian Thöni. 2016. "Hidden persuaders: Do small gifts lubricate business?" Working Paper.

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.

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*.

Offerman, Theo. 2002. "Hurting hurts more than helping helps." *European Economic Review*, 46(8): 1423–1437.

Pope, Devin G., Joseph P. Price, and Justin Wolfers. 2011. "Awareness reduces racial bias." NBER Working Paper Series No. 19765.

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

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

Salop, Steven C. 1979. "A model of the natural rate of unemployment." *American Economic Review*, 69(1): 117–125.

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

Shapiro, Carl, and Joseph E. Stiglitz. 1984. "Equilibrium unemployment as a worker discipline device." *American Economic Review*, 74(3): 433–44.

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.

Stiglitz, Joseph E. 1974. "Alternative theories of wage determination and unemployment in LDC's: The labor turnover model." *Quarterly Journal of Economics*, 88(2): 194–227.

Thakral, Neil, and Linh T. Tô. 2017. "Daily labor supply and adaptive reference points." Working Paper.

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.

Weiss, Andrew. 1980. "Job queues and layoffs in labor markets with flexible wages." *Journal of Political Economy*, 88(3): 526.

Yellen, Janet L. 1984. "Efficiency wage models of unemployment." *American Economic Review*, 74(2): 200–205.

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 $\tilde{e} \in \underset{e \in E}{\operatorname{argmax}}\, EU(e, \tilde{w}|e, \tilde{w})$. 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 $\tilde{e} \in \underset{e \in E}{\operatorname{argmax}}\, m(e, \tilde{w}) + n_e(e|\tilde{e}) \Rightarrow \tilde{e} = \underline{e}$. 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: $be - w_h = b\underline{e} - \underline{w} \Rightarrow e = (w_h - \underline{w})/b + \underline{e}$.

As shown in the text, the first two possibilities are given by:

$$e_g = \underline{e} + \frac{\alpha\eta K(w_h)b}{\gamma(1+\eta\lambda)} \qquad \text{requiring} \qquad w_h < \frac{\alpha\eta K(w_h)b^2}{\gamma(1+\eta\lambda)} + \underline{w} \tag{11}$$

and

$$e_l = \underline{e} + \frac{\alpha\eta\lambda K(w_h)b}{\gamma(1+\eta\lambda)} \qquad \text{requiring} \qquad w_h > \frac{\alpha\eta\lambda K(w_h)b^2}{\gamma(1+\eta\lambda)} + \underline{w}. \tag{12}$$

Figure 2 in the text 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 2, 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

As described in the text.

Proof of Proposition 2

Continuing from the proof of Corollary 1, and referring to Figure 2, 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^*.

As for part 2, existence of a profitable gift occurs when the condition stated in the proof of Proposition 2 holds, and this inequality trivially behaves as stated.

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

Proof of Proposition 3

Proceeding as in the proof of Corollary 1, as shown in the text, the possible optima that are not at the kink in the utility function, and the profit constraints that they require/imply, are given by

$$e_g = \underline{e} + \frac{\alpha\eta\lambda K(w_c) b}{\gamma(1+\eta\lambda)} \qquad \text{requiring} \qquad w_c - \underline{w} < \frac{\alpha\eta\lambda K(w_c) b^2}{\gamma(1+\eta\lambda)}. \tag{13}$$

and

$$e_l = \underline{e} + \frac{\alpha\eta\lambda^2 K(w_c) b}{\gamma(1+\eta\lambda)} \qquad \text{requiring} \qquad w_c - \underline{w} > \frac{\alpha\eta\lambda^2 K(w_c) b^2}{\gamma(1+\eta\lambda)} \tag{14}$$

Figure 5 is a more detailed version of Figure 3 that shows the curves defined by the RHS of these profit constraints, so that the inequalities hold with equality when $w_c = w_g$ and $w_c = w_l$ respectively, so that e_g is a valid local optimum when $w_c < w_g$ and e_l is a valid local optimum when $w_c > w_l$.

The worker must then compare these options, when they exist, to the kink in his utility function. The utilities of all three options are derived similarly to the positive reciprocity case. Comparing e_g or e_l to $\underline{e} + (w_c - \underline{w})/b$, we find that the utility at the kink is never optimal, similarly to the demonstration in the proof of Corollary 1; the difference between these propositions is that either e_g or e_l is always an available option in the negative surprise case, so that $\underline{e} + (w_c - \underline{w})/b$ is in fact never chosen. That is, $w_l < w_g$, to the contrary of Corollary 1.

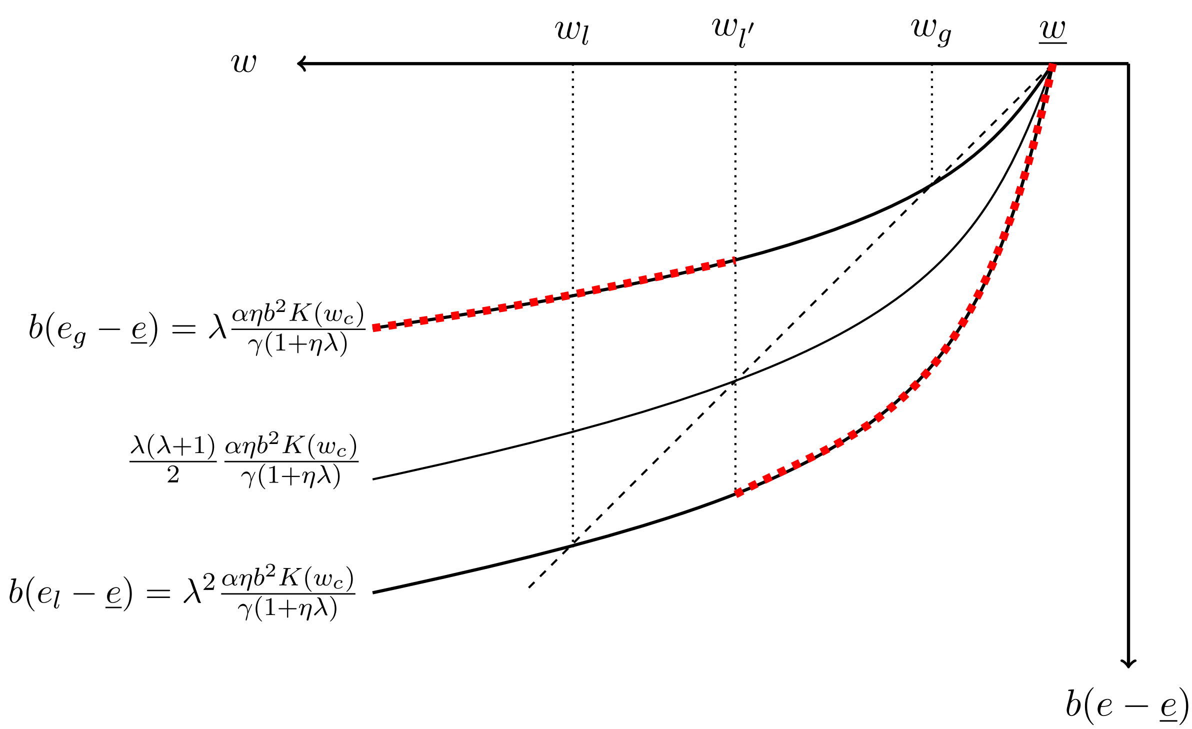
In the region between w_g and w_l where both e_l and e_g are valid optima, the worker prefers e_l to e_g only when

$$w_c - \underline{w} > \frac{\lambda(\lambda+1)}{2} \frac{\alpha\eta b^2 K(w_l)}{\gamma(1+\eta\lambda)}.$$

Define $w_{l'}$ as the value of w_c that makes this relationship hold with equality.

This relationship is a multiple of the revenue response curves that also determine the validity of the profit constraints above, so they are shown on Figure 5 together. By noticing that because $\lambda > 1$ and $1 < \frac{\lambda+1}{2} < \lambda$, the aggregate set of conditions imply that e_l is chosen for $w_c > w_{l'}$, and e_g is chosen otherwise. Regardless, a surprising wage cut is negatively reciprocated.

Figure 5: Revenue response function to fully surprising wage cut



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

Proof of Corollary 3

As described in text.

Proof of Corollary 4

As described in text.

Proof of Lemma 2

This can be seen by comparing the utility in the high wage state, given any effort plan, of $e_1 \leq \underline{e}$ versus $e_2 = \underline{e}+\epsilon$ for some small $\epsilon > 0$: If e_1 is strictly less than $\underline{e}$, then for sufficiently small ϵ both absolute effort costs and gain-loss effort costs will be lower when choosing e_2, and gain-loss reciprocity will also be higher when choosing e_2 because kindness is in the gain domain and e_2 leads to higher profits for the firm than e_1. Therefore, e_2 is strictly preferred to e_1 regardless of the effort plan, so the only credible plan has $\tilde{e}_h > \underline{e}$. Similarly, $\tilde{e}_l < \underline{e}$.

Proof of Lemma 3

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.

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 Lemma 2 and Assumption 2. 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.

Proof of Lemma 4

As in the 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

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

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

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

where both expressions rely on Lemma 2. 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 by Lemma 2 and the assumption that $w_l \geq \underline{w}$, 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 $e_h > \underline{e}$ by Lemma 2, this only holds if $p > 1/2$, contradicting the requirement that $p < 1/2$ for the worker to respond with these effort levels.

Proof of Proposition 4

Referring back to the proof of Lemma 3, the marginal utility functions for the worker upon realizing the wage state 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 anywhere where μ' is defined. Lemmas 3 and 4 prove that the PPE can never occur at the kinks in these utility functions where $\pi(e_l, w_l) = \pi(e_h, w_h)$ or $c(e_h) = c(e_l)$, but by the definition of a PPE, it *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:

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

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

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

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

which reduce to the following two constraints on the equilibrium effort and 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 with 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, by Lemma 2. 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

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

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

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

$$\begin{array}{rrcl} & p(b\tilde{e}_h - w_h) + (1-p)(b\tilde{e}_l - w_l) & > & b\underline{e} - \underline{w} \\ \Rightarrow & \frac{\eta\alpha b^2 \Delta K(1-p)p}{\gamma(1+\eta\lambda(1-p)+\eta p)} - \frac{\eta\lambda^2\alpha b^2 \Delta K(1-p)p}{\gamma(1+\eta p+\eta\lambda(1-p))} - pw_h - (1-p)w_l & > & -\underline{w} \\ \Rightarrow & \frac{\eta\alpha b^2 \Delta K(1-p)p}{\gamma}\left(\frac{1-\lambda^2}{1+\eta\lambda(1-p)+\eta p}\right) & > & pw_h + (1-p)w_l - \underline{w} \end{array} \tag{15}$$

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

$$\begin{aligned} p(b\tilde{e}_h - w_h) + (1-p)(b\tilde{e}_l - w_l) &> b\underline{e} - \underline{w} \\ \Rightarrow \quad \tfrac{\eta\alpha b^2\Delta K(1-p)p}{\gamma(1+\eta)} - \tfrac{\eta\lambda^2\alpha b^2\Delta K(1-p)p}{\gamma(1+\eta\lambda)} - pw_h - (1-p)w_l &> -\underline{w} \\ \Rightarrow \quad \tfrac{\eta\alpha b^2\Delta K(1-p)p}{\gamma}\left(\tfrac{1}{1+\eta} - \tfrac{\lambda^2}{1+\eta\lambda}\right) &> pw_h + (1-p)w_l - \underline{w} \end{aligned} \tag{16}$$

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

$$\begin{aligned} \tfrac{1}{1+\eta} &> \tfrac{\lambda^2}{1+\eta\lambda} \\ \Rightarrow \quad 0 &> \lambda^2(\eta+1) - \eta\lambda - 1 \end{aligned}$$

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 Corollary 5

As described in text.

Proof of Proposition 5

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 3, 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 3, 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 3 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.

B Appendix: Non-reference-dependent reciprocity

Incorporating baseline (non-reference-dependent) reciprocity into the model leads to the following utility function for workers:

$$u(e, w|\tilde{e}, \tilde{w}) = w - c(e) + \alpha K(w)\pi(e, w) + \mu(c(\tilde{e}) - c(e)) + \alpha\mu(K(w) - K(\tilde{w}))\mu(\pi(e, w) - \pi(\tilde{e}, \tilde{w})).$$

This utility function implies that workers generally reciprocate higher wages with higher effort. But it also affects the reciprocal response to surprising wage changes. This response is now moderated by the expected wage in addition to the gift wage and reservation wage, because workers may already be planning to exert above-minimal effort in order to reciprocate (positively or negatively) any expected wage that departs from the reservation wage.

Assuming that workers are hired at their reservation wage $\underline{w}$ and then surprised with a higher wage w_h thus leads to a greater reciprocal response than what is found in our primary analysis (see below). But if higher wages are profitable for the firm even without surprises, it makes sense for the firm to choose a higher wage from the beginning. This higher wage would thus be incorporated into the workers' reference points, and additional surprising wage changes would lead to positive or negative reciprocity relative to a baseline effort level that is higher than $\underline{e}$. As shown below, this makes positive surprising gifts less likely to exist, to the extent that a firm who is fully exploiting its workers' baseline reciprocity has *no* further room for profitable surprising gifts. It also greatly amplifies the dangers of failing to manage workers' expectations about future gifts. Our assumption of baseline reciprocity thus serves not only to emphasize the role of surprises, but represents a relatively agnostic and generous stance regarding the potential for profitable gift exchange.

First consider the response to a surprising wage w_h paid to workers who are expecting $\underline{w}$ with certainty. This is exactly the situation considered in Corollary 1, but with baseline reciprocity.

As in Corollary 1, upon realizing the high wage, workers maximize $U(e, w|\underline{e}, \underline{w})$. The first-order condition implies that

$$e = \underline{e} + \frac{\alpha b K(w_h)(1 + \eta\mu'_\pi)}{\gamma(1 + \eta\lambda)}$$

(so long as μ'_π exists). This gift is profitable when w_h is sufficiently small, in which case $\mu'_\pi = 1$ and $e = \underline{e} + \frac{\alpha b K(w_h)(1+\eta)}{\gamma(1+\eta\lambda)}$. Compared to the effort response in Corollary 1, $e = \underline{e} + \frac{\alpha b K(w_h)}{\gamma(1+\eta\lambda)}$, this is a larger degree of positive reciprocity, as expected.

However, once this high wage w_h has been incorporated into expectations, effort does not wane back to the minimal level $\underline{e}$. Instead, workers maximize their non-reference-dependent utility, which requires $c'(e) = \alpha b K(w_h)$ or $e = \frac{\alpha b}{\gamma} K(w_h) + \underline{e}$. The reader can verify that the optimal surprising wage raise is larger than the optimal wage once expectations have updated, and that any time there exists a profitable fully-surprising wage raise there is also a profitable fully-expected wage greater than the reservation wage. This suggests that any firm who wishes to use surprising gifts will not be starting from the reservation wage at all.

Following this reasoning, let's assume that the firm starts with a contracted wage at the optimal level, taking into account baseline reciprocity. Workers who anticipates a wage $\tilde{w} > \underline{w}$ maximize their non-reference-dependent utility, yielding $\tilde{e} = \underline{e} + \frac{\alpha b}{\gamma} K(\tilde{w})$ as above. The firm, anticipating this, maximizes its profits $\frac{\alpha b^2}{\gamma} K(\tilde{w}) - \tilde{w}$, yielding $K'(\tilde{w}) = \frac{\gamma}{\alpha b^2}$. With this as the expected wage, the firm then surprises the worker with a gift wage $w_h > \tilde{w}$. We can show that there is never any such wage that can lead the firm to higher profits, even in a one-shot interaction.

To see this, assume towards contradiction that such a wage $w_h > \tilde{w} = K'^{-1}\left(\frac{\gamma}{\alpha b^2}\right)$ exists that workers will respond to in such a way as to increase the firm's profits beyond what they would earn by sticking with the expected wage $\tilde{w}$. The worker responds to this surprise by choosing a new level of effort e_g satisfying the first order condition $\gamma(1+\eta\mu'_c)(e_g - \underline{e}) = \alpha b K(w_h)(1+\eta\mu'_\pi) - \alpha\eta b\mu'_\pi K(\tilde{w})$, where following the notation in the proof of Corollary 1, $\mu'_c = \mu'(c(\tilde{e}) - c(e_g))$. As opposed to that result, it is now possible for the worker to choose a level of effort with a lower cost than planned. Two kinks in the utility function exist, corresponding to the two dimensions of reference dependence.[34]

Because we are assuming that $\pi(e_g, w_h) > \pi(\tilde{e}, \tilde{w})$, and this can only happen if the worker increases effort (and the cost of effort) relative to his plan, $\mu'_c = \lambda$ and $\mu'_\pi = 1$. We can therefore define the function $e_g(w_h) = \underline{e} + \frac{\alpha b((1+\eta)K(w_h) - \eta K(\tilde{w}))}{\gamma(1+\eta\lambda)}$ that represents the effort response to the wage w_h *given* that this effort will in fact increase the firm's profits.

Now if we imagine a graph akin to Figure 2, plotting the revenue response $r(w_h) = b(e_g(w_h) - \tilde{e})$ as a function of the wage w_h, the curve $r(w_h)$ *must* rise above the 45-degree line $w_h - \tilde{w}$ in order for this effort response to successfully raise the firm's profits. Because of the concavity of K, this is equivalent to requiring that $\frac{\partial}{\partial w_h} r(w_h)|_{w_h=\tilde{w}} > 1$. But using the form of $e_g(w_h)$ above, this quantity equals $\frac{\alpha b^2(1+\eta)K'(w_h)}{\gamma(1+\eta\lambda)}|_{w_h=\tilde{w}} = \frac{\alpha b^2(1+\eta)\left(\frac{\gamma}{\alpha b^2}\right)}{\gamma(1+\eta\lambda)} = \frac{1+\eta}{1+\eta\lambda}$, which is always strictly less than 1. We have therefore contradicted our assumption that $e_g(w_h)$ does improve the firm's profits, for any possible $w_h > \tilde{w}$.

Our interpretation of these results is that the potential for profitable surprising gifts is mitigated by baseline reciprocity. This can be elaborated on further by examining the response to wage cuts relative to an above-market wage, which is intuitively strongly negatively reciprocated because it is a very cheap punishment to withdraw high effort. Combining this effect with the mitigated positive response to positive surprises, it is also intuitive that the potential for repeated gifts to be profitable is also lessened relative to the model analyzed in our primary results. Our analysis thus paints a perhaps overly-pessimistic view of reciprocity overall, but presents the most optimistic possible case for surprising gifts.

[34] Technically, three kinks exist because there are two points, symmetric around $\underline{e}$, at which effort costs cross from the gain domain to the loss domain. But it is trivially seen that in this case the worker would never consider effort below the lower kink, so we will ignore this region of the utility function.