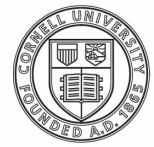


1-9-2015

A Theory of Fairness in Labor Markets

Daniel J. Benjamin
Cornell University, dl468@cornell.edu

A Theory of Fairness in Labor Markets

Abstract

I study a gift-exchange game, in which a profit-maximizing firm offers a wage to a fair-minded worker, who then chooses how much effort to exert. The worker judges a transaction fairer to the extent that his own gain is more nearly equal to the firm's gain. The worker calculates both players' gains relative to what they would have gained from the "reference transaction," which is the transaction that the worker most recently personally experienced. The model explains several empirical regularities: rent sharing, persistence of a worker's entry wage at a firm, insensitivity of an incumbent worker's wage to market conditions, and—if the worker is loss averse and the reference wage is nominal—downward nominal wage rigidity. The model also makes a number of novel predictions. Whether the equilibrium is efficient depends on which notion of efficiency is used in the presence of the worker's fairness concern, and which is appropriate to use partly depends on whether loss aversion is treated as legitimate for normative purposes.

Keywords

fairness, reference-dependence, gift exchange, rent sharing, money illusion, downward nominal wage rigidity

Disciplines

Labor Economics | Labor Relations

Comments

A Theory of Fairness in Labor Markets

Daniel J. Benjamin*
Cornell University, University of Southern California, and NBER

January 9, 2015

Abstract

I study a gift-exchange game, in which a profit-maximizing firm offers a wage to a fair-minded worker, who then chooses how much effort to exert. The worker judges a transaction fairer to the extent that his own gain is more nearly equal to the firm's gain. The worker calculates both players' gains relative to what they would have gained from the "reference transaction," which is the transaction that the worker most recently personally experienced. The model explains several empirical regularities: rent sharing, persistence of a worker's entry wage at a firm, insensitivity of an incumbent worker's wage to market conditions, and—if the worker is loss averse and the reference wage is nominal—downward nominal wage rigidity. The model also makes a number of novel predictions. Whether the equilibrium is efficient depends on which notion of efficiency is used in the presence of the worker's fairness concern, and which is appropriate to use partly depends on whether loss aversion is treated as legitimate for normative purposes.

JEL classification: D63, J31, M50

Keywords: fairness, reference-dependence, gift exchange, rent sharing, money illusion, downward nominal wage rigidity

*I am grateful to Philippe Aghion, Attila Ambrus, Antonia Atanassova, Gary Becker, Lynn Benjamin, James Choi, Steve Coate, Noam Elkies, Constança Esteves-Sorenson, Erik Eyster, John Friedman, Roland Fryer, Drew Fudenberg, Alexander Gelber, Ed Glaeser, Jerry Green, Oliver Hart, Ori Heffetz, Holger Herz, Ben Ho, Daniel Hojman, Richard Holden, Caroline Hoxby, Erzo Luttmer, Lisa Kahn, Lauren Kaiser, Emir Kamenica, Lawrence Katz, Miles Kimball, Nobuhiro Kiyotaki, Fuhito Kojima, Ilyana Kuziemko, David Laibson, Sendhil Mullainathan, Karthik Muralidharan, Emi Nakamura, Ted O'Donoghue, Masao Ogaki, Emily Oster, Stefan Penczynski, Herakles Polemarchakis, Giacomo Ponzetto, Alex Rees-Jones, Josh Schwartzstein, Jesse Shapiro, Andrei Shleifer, Joel Sobel, Jón Steinsson, Dmitry Taubinsky, Jeremy Tobacman, Stephen Weinberg, Richard Zeckhauser, and especially Edward Glaeser, David Laibson, Matthew Rabin, and Andrei Shleifer, as well as seminar participants at Harvard University, Cornell University, the University of Maryland, and the SAET Normative Economics Satellite Conference for valuable comments and advice. I thank the Program on Negotiation at Harvard Law School; the Harvard University Economics Department; the Chiles Foundation; the Federal Reserve Bank of Boston; the Institute for Quantitative Social Science; Harvard's Center for Justice, Welfare, and Economics; the National Institute of Aging, through Grant Number T32-AG00186 to the National Bureau of Economic Research; the Institute for Humane Studies; and the National Science Foundation for financial support. I am grateful to Samantha Cunningham, Julia Galef, Yuezhou Huo, Jelena Veljic, and Jeffrey Yip for excellent research assistance, and especially Hongyi Li and Derek Lougee, who not only provided outstanding research assistance but also made substantive suggestions that improved the paper. All mistakes are my fault. E-mail: daniel.benjamin@gmail.com.

1 Introduction

There is much evidence that workers' concern for "fair" transactions influences their labor market behavior. For example, Kahneman, Knetsch, and Thaler (1986) suggest that fairness explains rent sharing and "internal labor markets," the facts that workers' wages are relatively more sensitive to a firm's profits and relatively less sensitive to current labor market conditions than neoclassical theory might suggest. Bewley (1999) concludes that workers' feelings about fairness could explain why firms typically lay off workers rather than reduce wages: still-employed workers would consider wage cuts unfair and become less productive. Fehr, Goette, and Zehnder (2009) review these and other empirical findings and make the case that fairness concerns play an important role in labor markets.

This paper makes three contributions. First, building on existing models of fairness concerns (Fehr and Schmidt 1999, Charness and Rabin 2002), I develop a model of a worker's concern for fairness when interacting with a firm. A crucial element of the model is that the worker judges fairness by contrasting the current transaction with a "reference transaction," which is determined by the worker's recent personal experience. Second, I apply the model in a simple gift-exchange game and show that it can explain several labor market regularities: rent sharing, persistence of a worker's entry wage at a firm, insensitivity of an incumbent worker's wage to market conditions, and—with the additional assumptions that the worker is loss averse and evaluates losses with respect to his nominal wage—downward nominal wage rigidity. While many of these phenomena have explanations based on repetition or reputation, the model predicts they would continue to be observed in settings where repetition and reputation forces are weak. The model also makes some novel predictions, such as that effort will be upward rigid. Third, I analyze the efficiency of the equilibrium under alternative assumptions about whether fairness concerns and loss aversion are part of the worker's "true" preferences that are relevant for normative analysis.

Section 2 introduces the game that I study throughout the paper. The firm, which aims to maximize profit, offers a wage to each worker. Each worker then chooses how much effort to exert. To focus on the implications of the workers' fairness preferences, I assume that contracting is infeasible and the exchange is one-shot. Thus, if a worker were purely self-regarding, then the worker would exert minimal effort regardless of the wage, so in equilibrium the firm would not hire the worker.

Section 2 also develops the model of fairness concerns. It is an extension of commonly used specifications of preferences used to explain behavior in laboratory experiments (Fehr and Schmidt

1999, Charness and Rabin 2002). To allow the model to be applied to the transaction between a worker and a firm, I generalize the specification of preferences in two ways. First, I assume that the worker judges the fairness of a transaction not only with regard to the monetary transfer (i.e., the wage) but also with regard to the amount of effort exerted by the worker. That is, the worker judges the fairness of the transaction by comparing the gain in profit to the firm with the overall gain to the worker net of effort costs. Second, the actual transaction is contrasted with the *reference transaction*, a benchmark terms of exchange against which the worker views alternative transactions. The worker calculates his own and the firm's "surplus payoff" as the deviation of the player's actual payoff from the payoff he or she would have earned from the reference transaction. If the firm's surplus payoff and the worker's surplus payoff are equal, then the transaction is considered maximally fair. In contrast, the transaction is judged particularly unfair if one party's surplus payoff is much larger than the other's. The model captures the essential features of Kahneman, Knetsch, and Thaler's (1986) "dual entitlement theory." Consistent with some of Kahneman, Knetsch, and Thaler's (1986) survey data and with the experimental evidence in Herz and Taubinsky (2013), I assume that the worker's reference transaction (the wage and effort combination that determines the reference payoff) is determined by what the worker himself has recently personally experienced.

In Section 3, I study the equilibrium of this game, assuming that the worker has a strong concern for fairness. Because the worker is motivated by fairness, he chooses effort to equate the two players' surplus payoffs. Consequently, the worker is willing to exert more effort in response to a higher wage: a higher wage increases the worker's surplus payoff and reduces profit, so equating the surplus payoffs requires the worker to increase effort. In equilibrium, the firm offers the wage that induces the worker to exert the efficient level of effort. The reason is that, since the worker's effort choice will ensure that the players' surplus payoffs are equal, the firm maximizes its own surplus payoff by maximizing the sum of the surpluses.

The main empirical implication of the model in Section 3 is rent sharing: firms that are more profitable for a given level of the worker's effort—due to a higher output price or greater productivity—offer higher wages. In equilibrium a more profitable firm will offer a higher wage in order to induce the now-higher efficient level of effort. This implication is consistent with much evidence that more profitable firms pay higher wages to apparently identical workers (e.g., Abowd, Kramarz, and Margolis 1999) and more profitable industries pay higher wages to all occupations (e.g., Dickens and Katz 1987).

In Section 4, I examine a two-period, repeated version of the game in order to investigate the dynamic implications of the worker's fairness concerns. In this analysis, a key role is played by

the assumption that the transaction that takes place in period 1 becomes the worker's reference transaction for period 2.

Two main implications come out of the two-period model. First, workers who are paid more in period 1 (because they entered period 1 with a more favorable reference transaction) also end up getting paid more in period 2. That is because the higher pay in period 1 means that they enter period 2 with a more favorable reference transaction. Since the worker's effort choice equates the players' surpluses (relative to the reference transaction), the firm needs to offer a higher wage in order to induce the efficient level of effort. It is indeed an important empirical regularity that cohorts of workers who experience high entry wages continue to earn relatively high wages throughout their tenure at the firm (e.g., Baker, Gibbs, and Holmstrom 1994, Kahn 2010).

Second, the wage of the worker who remains employed by the firm in period 2 is insensitive to small variations in the worker's outside-option payoff. That is because the wage is entirely pinned down by the worker's reference transaction and the efficient level of effort. Both of these are independent of the worker's contemporaneous outside-option payoff. The empirical observation that incumbent workers' wages are determined in an "internal labor market" (internal to the firm) and largely shielded by fluctuations in external labor market conditions has been an important theme in the personnel economics literature (Doeringer and Piore 1971, Baker, Gibbs, and Holmstrom 1994, Seltzer and Merrett 2000).

Sections 5 and 6 extend the model to discuss downward nominal wage rigidity (DNWR), the fact that firms often avoid nominal wage cuts—choosing to freeze wages instead—but do not avoid nominal wage increases (e.g., Dickens et al 2007). Section 5 extends the model by assuming that the worker's fairness concerns exhibit "loss aversion": the worker judges a transaction as especially unfair if, relative to the reference transaction, the worker receives a lower wage or exerts higher effort. Given this assumption, the worker's effort is more responsive to wage cuts than wage increases. As a result, when faced with a range of shocks to its output price, the firm optimally freezes the wage rather than cutting it. Sections 6 adds the additional assumption that the monetary amounts in the worker's reference transaction are nominal quantities, rather than real quantities. Besides providing a formal model of DNWR, the analysis makes a variety of novel predictions regarding how wage and effort respond to shocks to the firm's output price and how these effects vary depending on whether the economic environment is characterized by generally increasing, decreasing, or stable prices.

Section 7 addresses the efficiency of the equilibrium transaction. There are several possible generalizations of Pareto efficiency that can be applied, depending on whether efficiency is judged

in terms of the purely self-regarding component of the worker's payoff or in terms of the utility function that represents the worker's behavior, which includes fair-mindedness and possibly also loss aversion. Which notion is normatively appropriate depends on what the worker's "true" preferences are, by which I mean what the worker would choose with accurate beliefs and after deliberation. If the worker is fair-minded but not loss averse, then it does not matter which efficiency criterion is used because the equilibrium transaction is efficient according to both notions. However, if the worker is both fair-minded and loss averse, then the equilibrium may not be efficient in terms of utility and generally is not efficient in terms of the purely self-regarding component of preferences.

Section 8 mentions other contexts outside the labor market for which the fairness model developed in this paper may yield useful insights. The focus of the section, however, is on directions in which the model might be extended to be more realistic. Two directions merit discussion here (rather than in Section 8) because there has already been much closely related work in the behavioral economics literature.

One direction is to explore alternatives to my assumption that the worker's reference transaction is wholly determined by the worker's recent personal experience. This assumption plays a key role in enabling the model to capture empirical regularities regarding wage changes. However, there are also other plausible reference transactions that may matter in some settings. As in other contexts of reference-dependent preferences, the reference point is likely to be at least partly influenced by expectations (Köszegi and Rabin 2006); see Esteves-Sorenson, Macera, and Broce (2014) and Eliaz and Spiegler (2014) for models of fairness with an expectation-based reference point. Moreover, in labor market contexts, much work has emphasized workers judging fairness by comparing their own wage and effort with that of other workers. Akerlof and Yellen (1990) argue that such social comparisons may explain jealousy between workers, wage compression within firms, wage secrecy norms, and the negative correlation between occupational skill and unemployment. While the most direct tests from laboratory experiments find little evidence that workers' behavior is sensitive to how much other workers are paid (Maximiano, Sloof, and Sonnemans 2007, Charness and Kuhn 2007), field evidence indicates that such social comparisons influence job satisfaction and may affect turnover (Card, Mas, Moretti, and Saez 2012). Finally, if firms have some ability to shape their workers' reference transactions, then they would have an incentive to do so.

The second direction is to incorporate important aspects of fairness preferences that are omitted from the model, such as reciprocity (e.g., Rabin 1993) or social-image or self-image concerns (e.g., Andreoni and Bernheim 2009).

The most closely related paper is Benjamin (2014), a companion paper that studies the same

basic setup with a more general class of preferences. The present paper focuses on drawing out implications of workers' fairness concerns for empirical labor market regularities. To that end, here I formally incorporate the reference transaction into the model, which is important for studying the implications of fairness concerns in a two-period model, and I incorporate loss aversion, which is important for studying wage rigidity. Benjamin (2014) and the present paper jointly supplant my earlier working paper, Benjamin (2005). Fehr, Goette, and Zehnder (2009) provide an overview of how workers' fairness concerns relate to empirical evidence from labor markets and provide intuition very much in line with the formal model I develop here. Eliaz and Spiegler (2014) develop a formal model that addresses some of the same empirical evidence. Their approach is complementary with the present paper's since they embed the firm-worker relationship into a matching model of labor market equilibrium but model the worker's fairness concerns in a more reduced-form way.

2 Model Setup

2.1 The gift-exchange game

To focus on the basic workings of the model, I begin by analyzing a single-period game. There is a firm and a large number N of identical workers. The firm simultaneously chooses each worker i's salary, $w_i \in \mathbb{R}$, which I refer to as the "wage." (In principle, the firm could make the wage contingent on other variables, but as I explain at the beginning of Section 3 below, the firm will not be able to do better than an uncontingent wage.) Then each worker simultaneously chooses his level of effort, $e_i \in \mathbb{R}$. I assume that effort is observable but not verifiable.

Before the game begins, the firm could choose not to hire a given worker, or the worker could choose not to work. In that case, the firm gets zero effort from that worker and pays zero wage to him, and the worker earns outside-option utility zero. As a tie-breaker, I assume that the firm and worker choose employment if indifferent with the outside option.

For simplicity, I assume that the firm's production function is linear in effort, and effort is the only input. Thus, the firm's **total profit** is $\Pi \equiv p\sum_{i=1}^{N} e_i - \sum_{i=1}^{N} w_i$. I refer to the exogenous parameter p as the price of the firm's output, but it can also represent the productivity of the firm or workers. The firm's total profit is verifiable, but since no individual worker's effort is verifiable, the firm's **profit from worker** i, $\pi_i(w_i, e_i; p) = pe_i - w_i$, is not verifiable.

Each worker i's **material payoff** is $u_i(w_i, e_i) = w_i - c(e_i)$, where $c(e_i)$ is the worker's cost-of-effort function satisfying $c(0) = 0$, $c' > 0$, $c'' > 0$, $c'(0) < 1$, and $\lim_{e_i \to \infty} c'(e_i) = \infty$. Note that since the material payoff function is quasi-linear in the wage, the cost of effort and the material

payoff are denominated in monetary units.

The firm's objective is to maximize profit. In contrast, a worker's material payoff represents the purely self-regarding component of his outcome from the transaction but not necessarily the utility function that his behavior maximizes. A worker's utility when employed, denoted U_i^E, may depend on both his material payoff u_i and the firm's profit from interacting with him π_i; the worker's utility function is discussed below.

Everything is common knowledge.[1] The equilibrium concept is subgame-perfect Nash equilibrium.

Since workers are identical and their bilateral interactions with the firm are independent, the equilibrium for each bilateral interaction will be identical. Therefore hereafter, I refer to π_i as simply "profit," and I drop the i subscripts from all variables to reduce notational clutter. I call the outcome of the game, (w, e), a **transaction**.

The **efficient level of effort**, denoted $e_{\text{eff}}(p)$, is defined by $p = c'(e_{\text{eff}})$. In Section 7, I discuss efficiency in greater detail; here I merely remark that $e_{\text{eff}}(p)$ is the effort that maximizes the "material gains from trade" from the transaction, defined as the sum of the firm's profit $\pi(w, e; p)$ and the worker's material payoff $u(w, e)$. Since the wage is merely a transfer between the firm and the worker, the material gains from trade does not depend on it. I denote the material gains from trade at the efficient effort $e_{\text{eff}}(p)$ by $M(p) \equiv \pi(w, e_{\text{eff}}(p); p) + u(w, e_{\text{eff}}(p)) = pe_{\text{eff}}(p) - c(e_{\text{eff}}(p))$. In the analysis below, I assume that $p > 1$ so that both $e_{\text{eff}}(p)$ and $M(p)$ are positive.

2.2 The reference transaction and concern for fairness

Several models of fairness concerns have been proposed—such as those by Fehr and Schmidt (1999) and Charness and Rabin (2002)—to describe how people trade off their own material payoff against others' material payoffs. However, these models cannot be used naïvely as the specification for the worker's utility because they are tailored to behavior in laboratory settings. In particular, the models specify utility over the domain of the experimental participants' monetary gains or losses from the experiment. To apply these models to study a labor market interaction, two generalizations are needed.

First, in order to capture the players' overall gains or losses from the labor market transaction,

[1] If the firm were uncertain about whether the worker has fairness concerns or is purely self-regarding, then the equilibrium wage would be lower. Intuitively, by offering a lower wage, the firm can get some of the benefit if the worker turns out to be fair-minded while insuring against losing too much if the worker turns out to be purely self-regarding. Since the wage would be lower, the equilibrium effort exerted by a fair-minded worker would not be efficient (as in Fehr and Schmidt's (1999) analysis of the gift-exchange game). The normative conclusions in Section 7 would have to be modified, but the comparative statics in Sections 3-6 would have the same signs.

the utility function must take into account not only money but also effort. Therefore, in the formulation that follows, the worker's utility will depend on the firm's profit (from interacting with that worker) and the worker's material payoff, which are functions of both monetary payment and effort. This formulation specializes to the existing models in a laboratory environment, in which effort e is a number chosen by an experimental participant (instead of being real effort) and the payoffs π and u are monetary amounts paid to the participants.

Second, the utility function must take into account that fairness is judged relative to a "reference transaction." This phenomenon is clearly illustrated in Kahneman, Knetsch, and Thaler's (1986) evidence. Their data indicate that survey respondents consider transactions that adhere to recently experienced terms of exchange to be fair, even though the transactions do not equalize the agents' gains. For example, they find that people consider it unfair for a landlord to raise rents on existing tenants, yet fair to charge a new tenant a higher price when the old tenant leaves. Most relevantly to the current setting, when the market wage falls, respondents consider it unfair for a firm to reduce a current worker's wage to the going wage but fair to hire a new worker at that rate. Based on evidence from these and other scenarios, Kahneman, Knetsch, and Thaler proposed that an individual perceives a transaction as unfair if it deviates from the "reference transaction," which they describe as recent past experience, aspirations, or the going market terms of employment. The laboratory-based models can be viewed as a special case in which the reference transaction is that all participants in the experiment have zero earnings.

I formalize the **reference transaction**, $(w_0, e_0; p_0)$, as a particular transaction (wage and effort) occurring at a particular value of the output price. I refer to the payoff a player would get from the reference transaction as the player's **reference payoff**: the firm's reference profit is $\pi_0 \equiv \pi(w_0, e_0; p_0)$, and the worker's reference material payoff is $u_0 \equiv u(w_0, e_0)$.

The perceived fairness of a transaction depends on how players' payoffs, relative to their reference payoffs, compare to each other. To capture this idea, I define the **firm's surplus** from transaction (w, e) occurring at price p as $\widetilde{\pi}(w, e; p) \equiv \pi(w, e; p) - \pi_0$ and the **worker's surplus** as $\widetilde{u}(w, e) \equiv u(w, e) - u_0$. (For the functions $\widetilde{\pi}$ and $\widetilde{u}$ and some others below, I suppress dependence on the reference transaction for notational compactness.) The fairness function $f(\widetilde{u}, \widetilde{\pi})$ describes the worker's judgment about the fairness of his own transaction with the firm and is discussed further below.

The worker's utility function is

$$U = \begin{cases} U^E \equiv \sigma u + (1-\sigma) f(\widetilde{u}, \widetilde{\pi}) & \text{if the worker is employed} \\ 0 & \text{if the worker is not employed} \end{cases} . \tag{1}$$

The worker's utility when employed is the weighted sum of a self-interested component and a fairness component, where $0 < \sigma < 1$ is a preference parameter describing how much the worker cares about himself relative to fairness. The classical model of pure self-interest is the case $\sigma = 1$.

2.3 The fairness function

I assume that the fairness function is piecewise linear:

$$f(\widetilde{u}, \widetilde{\pi}) = \begin{cases} \rho_A \widetilde{u} + (1-\rho_A)\widetilde{\pi} & \text{if } \widetilde{u} > \widetilde{\pi} \\ \rho_D \widetilde{u} + (1-\rho_D)\widetilde{\pi} & \text{if } \widetilde{u} \leq \widetilde{\pi} \end{cases} , \tag{2}$$

where $1 > \rho_A \geq 0$ is the relative weight on the worker's surplus in the case of **advantageous unfairness**, which is when the worker's surplus exceeds the firm's, and $\rho_D > \rho_A$ is the relative weight on the worker's surplus in the case of **disadvantageous unfairness**, when the firm's surplus exceeds the worker's.[2]

Given fairness function (2), the worker's utility when employed can be re-parameterized and written as:

$$U^E = \begin{cases} \mu_A \tilde{u} + (1-\mu_A)\tilde{\pi} + \sigma u_{\mathbf{0}} & \text{if } \widetilde{u} > \widetilde{\pi} \\ \mu_D \tilde{u} + (1-\mu_D)\tilde{\pi} + \sigma u_{\mathbf{0}} & \text{if } \widetilde{u} \leq \widetilde{\pi} \end{cases} , \tag{3}$$

where $\mu_A \equiv \sigma + \rho_A(1-\sigma)$ and $\mu_D \equiv \sigma + \rho_D(1-\sigma) > \mu_A$ are composite parameters describing the overall weight on the worker's surplus. The utility function (3) generalizes common specifications in the literature. When the surpluses are incremental monetary payoffs from an experiment (in which case $u_{\mathbf{0}} = \pi_{\mathbf{0}} = 0$), parameter values satisfying $\mu_D > 1 > \mu_A > 0$ corresponds to Fehr and Schmidt's (1999) inequity-aversion model, while Charness and Rabin (2002) argue that $1 > \mu_D > \mu_A > 0$.

3 The Single-Period Game

In this section, I analyze the model laid out in the previous section. First note that the setup of the game rules out motivating the worker with a contract. Because effort is unverifiable, the firm

[2] The linearity of the two parts of the function is a simplifying assumption. However, the assumptions that the fairness function is kinked and that the kinks occur at equal surpluses are substantive. The utility function U^E resulting from this fairness function are an example of "fairness-kinked preferences," discussed in greater generality in Benjamin (2014).

cannot make the wage contingent on effort. Since total profit Π is verifiable, the firm could make each worker's wage an increasing function of Π. But as is well known in such settings (Prendergast 1999, pp. 41-42), as long as the number of workers employed by the firm is large, the incentive effects would be negligible. For example, if the firm set each worker's wage equal to $\frac{\Pi}{N}$, then the worker's gain from increasing profit by \$1 would be only $\$\frac{1}{N}$. Also note that because the worker's cost-of-effort function is convex, it is strictly better for the firm to pay a certain wage than a wage that is contingent on a random variable. Thus, without loss of generality, we can consider the firm's strategy to be the choice of an uncontingent wage level.

If the worker were purely self-interested, then the players would not transact. To discuss this case, suppose that w_i and e_i are bounded below by finite values $\underline{\underline{w}} < 0$ and $\underline{\underline{e}} < 0$, respectively. Regardless of the wage, the worker would choose the lowest possible effort $\underline{\underline{e}}$ because doing so maximizes his material payoff. Knowing this, the firm would offer the lowest possible wage $\underline{\underline{w}}$. Thus, at least one of the players would prefer his outside option. In contrast, as is well known, if the worker has fairness concerns, then it may be possible to realize gains from trade (e.g., Fehr and Schmidt 1999).

3.1 The worker's effort choice

The reason trade can occur is that the worker's fairness concerns make his effort choice an increasing function of the wage, as long as the worker's fairness concerns are strong enough. The following assumption provides sufficient conditions on the parameter values:

Assumption A. *(i)* $\mu_D \geq 1$, *and (ii)* $\mu_A \leq \frac{1}{2}$.

In words, A(i) states that when the transaction is advantageously unfair, the worker puts negative weight on the firm's payoff and positive weight on his own. As noted above, such "behindness aversion" is one of the assumptions underlying Fehr and Schmidt's (1999) inequity-aversion model. There is debate over whether behindness aversion is a reasonable assumption, and most evidence from dictator-game experiments is inconsistent with it (for discussion, see Benjamin's (2014) footnote 6 and accompanying text). A(ii) states that the worker puts greater weight on the firm's payoff than on his own when the transaction is advantageously unfair. The estimates from Fehr and Schmidt (1999) and Charness and Rabin (2002) are both consistent with a sizeable minority of experimental participants having $\mu_A \leq \frac{1}{2}$. The role of each part of Assumption A and the scope for relaxing each part are discussed below after Lemma 1 and Proposition 1.

Let the worker's utility-maximizing effort when employed be denoted $e(w,p)$. Let $\tilde{e}(w,p)$, defined by $\widetilde{\pi}(w,\tilde{e};p) = \widetilde{u}(w,\tilde{e})$, denote the level of effort that equates the players' surpluses.

Lemma 1. *Under Assumption A, for any $p > 1$, there exists $\overline{w}(p)$ such that:*

1. *If $w \leq \overline{w}(p)$, then $e(w,p) \equiv \tilde{e}(w,p)$. Moreover, $\tilde{e}(w,p)$ is increasing in w and decreasing in p.*

2. *If $w > \overline{w}(p)$, then $e(w,p) \in [e_{\text{eff}}(p), \tilde{e}(w,p))$, and $e(w,p)$ is constant in w and increasing in p.*

All proofs are relegated to the appendix.

Part 1 of Lemma 1 states that as long as the wage is below some threshold $\overline{w}$, the worker chooses effort so as to equate the players' surpluses from the transaction. To understand why, note that whenever $e < e_{\text{eff}}$, a marginal increase in effort increases the firm's profit more than it reduces the worker's material payoff. Due to Assumption A(ii) ($\mu_A \leq \frac{1}{2}$), the worker when getting the majority of the surplus puts at least as much weight on the firm as himself. Due to Assumption A(i) ($\mu_D \geq 1$), the worker when earning less than half the surplus puts non-positive weight on the firm. Consequently, for any wage at which the worker ends up exerting less than the efficient level of effort, the worker would increase his effort exactly up to (and not beyond) the level that equates the surpluses.

Part 2 of the lemma states that there is a maximum level of effort that the worker is willing to exert, and this maximum level of effort is above the efficient level. At wages higher than the threshold $\overline{w}$, the worker's effort (equal to the maximum) would be lower than the equal-surplus level. However, Part 2 of the lemma will not be relevant for the equilibrium (discussed below) because the firm will never want to offer a wage higher than necessary for inducing the efficient level of effort.

If Assumption A(ii) were violated, then the threshold $\overline{w}$ would be low enough that the maximum level of effort that the worker is willing to exert would be below the efficient level. If Assumption A(i) were violated, then there would also be a *minimum* level of effort that the worker is willing to exert, and at wages below some threshold, the worker's effort (equal to the minimum) would be higher than the equal-surplus level. We discuss these assumptions further below in the context of the equilibrium.

Part 1 of Lemma 1 (the relevant part for the equilibrium) also states that effort is increasing in the wage, holding price constant. That is because a higher wage transfers surplus from the firm

to the worker, so equating the surpluses requires higher effort. There is evidence from police (Mas 2006) and airline pilots (Lee and Rupp 2007) that plausibly exogenous changes in the wage cause corresponding changes in performance. In laboratory labor markets with one-shot, anonymous interactions, experimental economists have consistently found that higher wage offers induce greater effort (e.g., Fehr, Kirchsteiger, and Riedl 1993, Fehr, Kirchsteiger, and Riedl 1998, Fehr and Falk 1999). There is an increasing number of experiments that study the effect of wage increases when subjects have been hired into a realistic job setting (for an early study along these lines, see Pritchard, Dunnette, and Jorgenson 1972). One study finds that effort increases when the wage increases (Cohn, Fehr, and Goette 2014), while several others find no effect (Hennig-Schmidt, Rockenbach, and Sadrieh 2010, Kube, Maréchal, and Puppe 2013, Esteves-Sorenson and Macera 2013). Gneezy and List (2006) find a positive effect that fades over the course of a few hours.

The final statement in Part 1 of Lemma 1 is that effort is decreasing in the price, holding the wage constant. That is because, all else equal, an increase in price increases the firm's surplus, so equating the surpluses requires decreasing effort. I am not aware of evidence regarding this prediction.

3.2 The equilibrium

Given the worker's effort function, the firm's choice of wage pins down the equilibrium transaction. The following assumption provides sufficient conditions on the reference payoffs u_0 and π_0 for the worker to be employed in equilibrium:

Assumption B. *(i)* $0 \leq u_0, \pi_0 \leq M(p)$, *and (ii)* $(1-2\sigma)\,u_0 + \pi_0 \leq M(p)$.

B(i) requires that neither player's reference payoff is too low or too high, and B(ii) requires that their weighted sum is not too high. Note that if the worker puts more weight on his material payoff than on fairness ($\sigma \geq \frac{1}{2}$), then B(ii) is redundant with B(i). I return to Assumption B and discuss it in more detail below.

Proposition 1 states that at the equilibrium transaction, the surpluses are equal and effort is efficient.

Proposition 1. *Under Assumptions A and B, for any* $p > 1$, *there is a unique equilibrium in which the firm hires the worker, and the equilibrium transaction* (w^*, e^*) *satisfies* $\pi(w^*, e^*; p) - \pi_0 = u(w^*, e^*) - u_0$ *and* $e^* = e_{\text{eff}}(p)$.

Benjamin's (2014) Proposition 2 is a closely related result (slightly less general in assuming $\pi_0 = u_0 = 0$ but slightly more general in allowing some values of μ_D less than 1). The logic of the equilibrium is straightforward. Lemma 1 showed that, faced with a given wage below the threshold $\overline{w}$, the worker chooses effort so as to equate the players' surpluses. Knowing this, the firm maximizes profit by offering the wage level that induces the worker to choose the efficient level of effort. Lemma 1 implies that the required wage for efficient effort is in fact below $\overline{w}$.

The role of Assumption B is to ensure that both the worker and the firm prefer the equilibrium to their outside options. For example, if the worker cares at least as much about himself as about fairness ($\sigma \geq \frac{1}{2}$), then even if $u_0 = M$ and $\pi_0 = M$—so that the worker will judge any transaction as unfair to at least one of the players—both players will still choose to transact at a wage that gives all the gains to trade to the worker: the firm earns zero profit, and $U^E \geq 0$ because the worker's gain in material payoff outweighs the disutility from unfairness. If the worker cares mostly about fairness ($\sigma < \frac{1}{2}$), Assumption B imposes an additional restriction on the reference payoffs because the scope for unfairness to be offset by a gain in the worker's material payoff is more limited. For example, in the extreme case in which the worker cares exclusively about fairness ($\sigma = 0$), Assumption B imposes the additional restriction that $u_0 + \pi_0 \leq M$; if this restriction is violated, then any transaction will be unfair to at least one of the players, and thus the worker will prefer his outside option.

The result that the worker's fairness concerns enable fully efficient exchange is perhaps surprising. In a more general model, Benjamin's (2014) Theorems 2 and 4 provide necessary and sufficient conditions, respectively, for this to occur. In the present context, Assumption A(ii) ($\mu_A \leq \frac{1}{2}$) plays a key role in enabling the equilibrium to be efficient. As noted after Lemma 1, if Assumption A(ii) were violated, then there would be a maximum level of effort that the worker were willing to exert that would be lower than the efficient level. However, as long as μ_A were not too much smaller than $\frac{1}{2}$, the equilibrium for this set of parameter values would—aside from not being efficient—nonetheless be qualitatively similar to the equilibrium in Proposition 1: for wages below the lowest level $\overline{w}$ that induces maximal effort, the worker would exert effort that ensures equal surpluses, and the firm would therefore maximize profit by offering wage $\overline{w}$. The comparative statics at this inefficient equilibrium would be the same as those described below in Proposition 2 and later throughout the paper. In contrast, if μ_A were too small, then the worker would be almost wholly self-interested, and the equilibrium outcome would be no trade.

Assumption A(i) ($\mu_D \geq 1$) could be relaxed somewhat without affecting the equilibrium. As noted after Lemma 1, if $\mu_D < 1$, the worker would be willing to exert effort higher than the equal-

surplus level at low enough wages. However, as long as μ_D were close enough to 1, the firm would still maximize profit by offering the higher wage w^*. In contrast, if μ_D were too small, then the worker would exert relatively high effort at a relatively low wage, and the firm could earn higher profit by exploiting the worker's generosity with a low wage.

At the equilibrium from Proposition 1, Proposition 2 outlines the comparative statics.

Proposition 2. *At the equilibrium described in Proposition 1:*

1. w^* *and* e^* *are both increasing in* p.
2. e^* *does not depend on* u_0 *nor* π_0.
3. w^* *is increasing in* u_0 *and decreasing in* π_0.

Part 1 states that the equilibrium wage and effort are increasing in the firm's output price p. When the price goes up, equilibrium effort is higher because the efficient level of effort is now higher. Lemma 1 states that effort would be reduced if the price increased with the wage held constant, but since the firm wants the worker to exert *more* effort, the firm must offer a higher wage in equilibrium.

Part 1 implies that the firm and worker share the rents when the firm becomes more profitable or productive. Such rent sharing is consistent with much evidence that more profitable firms pay higher wages to apparently identical workers (e.g., Abowd, Kramarz, and Margolis 1999) and more profitable industries pay higher wages to all occupations (e.g., Dickens and Katz 1987). Relatedly, many firms institutionalize the positive relationship between profit and wages by paying workers through profit-sharing plans, gain-sharing plans, or stock options (Kruse et al 2003). Seventy percent of firms with profit-sharing plans believe they improve productivity (Ehrenberg and Milkovich 1987), and there is evidence that this is true (Weitzman and Kruse 1990, Kruse 1993). The positive effect of profit-sharing on worker performance is a puzzle for standard incentive theory with self-interested workers because, as noted at the beginning of this section, free riding by workers makes the potential positive incentive effects negligible (Prendergast 1999).[3]

While more profitable firms may pay higher wages for a number of reasons—for example, to attract higher-ability workers—several sources of evidence indicate that rent sharing may be at least partly due to workers' fairness concerns. For one thing, managers themselves say that fairness

[3]Even if workers can monitor each other and punish poor performance, workers would be expected to free ride on monitoring in companies with many workers.

perceptions play a primary motivational role in real-world wage policies (e.g., Blinder and Choi 1990, Levine 1993, Agell and Lundborg 1995, Campbell and Kamlani 1997, Bewley 1999). In addition, rent sharing arises in anonymous, one-shot laboratory labor markets that rule out alternative mechanisms (Fehr, Kirchsteiger, and Reidl 1993, Fehr, Kirchsteiger, and Reidl 1998, Falk and Fehr 1999, Brown, Falk, and Fehr 2004).

In labor economics, rent sharing is often modeled as the outcome of Nash bargaining between a firm and a worker. While Nash bargaining also leads to the worker exerting efficient effort and the two players splitting the surplus, there is an important difference. In a bargaining model, the surpluses are calculated relative to the firm's and worker's outside options, whereas in the fairness model, the surpluses are calculated relative to the worker's reference transaction. Thus, the fairness model predicts that a worker with a more favorable reference transaction (say, due to having had a better deal in his last job) should be paid more than a worker with an identical outside option but who has a less favorable reference transaction.

Returning to Proposition 2, Parts 2 and 3 provide comparative statics with respect to changes in the reference transaction. The efficient level of effort does not depend on the reference transaction, and hence the equilibrium effort is independent of u_0 and π_0. However, at the efficient level of effort, when the reference transaction is more favorable to the worker (u_0 is greater) or less favorable to the firm (π_0 is smaller), equating the surpluses requires giving the worker a higher material payoff and the firm a lower profit. Therefore the equilibrium wage is higher.

4 The Two-Period Game

In the previous section, the reference transaction was treated as an exogenous constant. The key difference in the dynamic version of the model in this section is that the reference transaction evolves over time. I show that if the reference transaction is shaped by the worker's recent personal experience, then the model can explain two important empirical regularities about wage dynamics: (1) workers paid more at time of hire earn higher wages subsequently, and (2) the wage of a worker who remains at a firm is largely unaffected by variation in external labor market conditions.

To address these stylized facts in as simple a model as possible, I study a two-period setting. In period $t = 1$, the worker is "new" at the firm. The firm makes a take-it-or-leave-it wage offer w_1 to the worker. If the worker refuses, the game ends, and the players get their outside-option profit/utility of 0 in both periods. If the worker accepts, then the worker chooses effort e_1, and the game continues into period 2. The firm makes a take-it-or-leave-it wage offer w_2 to the "incumbent"

worker. The worker can refuse, in which case both players get their outside-option profit/utility of 0 for that period, or accept and choose effort e_2. Profit and the worker's material payoff in each period are the same as in the single-period game from the previous section. In period 1, the firm maximizes the expected sum of its profit in each period, $\pi_1 + E_1\pi_2$, and the worker maximizes $U_1 + E_1U_2$. The expectation appears because p_t is a random variable, which I assume is drawn i.i.d. from an atomless distribution that has full support on $(1, \infty)$. In both periods, as a tie-breaker, I assume that the players choose to transact if indifferent. In the pre-game "period 0," the worker was employed in the external labor market, not at the firm.

To complete the model, I assume that the reference transaction is the worker's recent personal experience: the period-1 reference transaction reflects the worker's experience prior to employment with the firm (in "period 0") and is taken as exogenous, and the period-2 reference transaction equals the transaction that occurred in period 1. The reference transaction thus links the two periods. The "recent-experience assumption" is consistent with some experimental evidence (Herz and Taubinsky 2013) as well as much survey evidence such as Kahneman, Knetsch, and Thaler's (1986) mentioned in Section 2.2 (see also Thaler 1985 and Bolton, Warlop, and Alba 2003).

For the two-period game, Assumptions A and B need to be modified as follows:

Assumption A′. *(i)* $\mu_D \geq 1$ *and (ii)* $\mu_A \leq \frac{1-\sigma}{2}$.

Assumption B′. *(i)* $0 \leq u_0, \pi_0 \leq \frac{M(p_1)+EM(p_2)}{2}$, *and (ii)* $\left(\frac{1-3\sigma}{\sigma+1}\right) u_0 + \pi_0 \leq \frac{\sigma M(p_1)+EM(p_2)}{\sigma+1}$.

A′(i) is identical to A(i), but A′(ii) imposes a tighter bound than A(ii). I discuss the role of A′(ii) after presenting the proposition below. Assumption B′ is analogous to Assumption B but differs because when deciding whether or not to transact in period 1, the players take into account the anticipated period-2 equilibrium payoffs. Relative to B′(i), B′(ii) imposes a non-redundant restriction whenever $\sigma < 1$.

The game is straightforward to solve using backward induction. Since period 2 is the single-period game, period-2 behavior is as described in the previous section. The period-1 transaction affects period 2 by becoming the worker's reference transaction. If the worker were purely self-interested, then the players would not transact in either period.

Proposition 3 characterizes the subgame-perfect equilibrium of the two-period game.

Proposition 3. *Under Assumptions A′ and B′, with positive probability there is a unique equilibrium in which the firm hires the worker in both periods. The equilibrium transactions,* (w_t^*, e_t^*) *for* $t = 1, 2$, *satisfy* $\pi(w_t^*, e_t^*; p_t) - \pi_{t-1} = u(w_t^*, e_t^*) - u_{t-1}$ *and* $e_t = e_{\text{eff}}(p_t)$.

As in Proposition 1, Assumption B′—that u_0 and π_0 are neither too low nor too high—helps ensure that the firm hires the worker in period 1. If the period-2 price realization is much lower than the period-1 price realization, however, then π_1 and u_1 may be so high that the worker prefers his outside option in period 2.

Assuming the price realizations make it profit-maximizing for the firm to hire the worker in both periods, Assumption A(ii) ($\mu_A \leq \frac{1}{2}$) is sufficient to imply that the worker chooses effort so as to equate the surpluses *in period 2*. In period 1, however, the worker anticipates that higher effort will lead to a less favorable reference transaction for period 2 and therefore a lower equilibrium period-2 material payoff. Since higher effort in period 1 reduces not only the worker's period-1 material payoff but also his period-2 material payoff, the maximum level of effort the worker is willing to exert is lower in period 1 than in period 2. The role of A′(ii) ($\mu_A \leq \frac{1-\sigma}{2}$) is to ensure that the period-1 maximum effort is nonetheless higher than the efficient level, and thus the equilibrium effort in period 1 is also efficient.[4]

Proposition 4 outlines the comparative statics.

Proposition 4. *At the equilibrium described in Proposition 3: for $t = 1, 2$,*

1. *w_t^* and e_t^* are both increasing in p_t.*
2. *e_t^* does not depend on u_0 nor π_0.*
3. *w_t^* is increasing in u_0 and decreasing in π_0.*

Part 1 says that each period's wage and effort is higher if the firm's output price in that period is higher. This is the same rent sharing as in Proposition 2. Just as in the single-period model, in each period the firm maximizes profit by inducing the efficient level of effort. Because the worker chooses effort to equate the surpluses, inducing higher effort when the output price is higher requires paying a higher wage. This result for the two-period model predicts that rent sharing should be observed not only in the cross section across firms but also in the time series within a firm.

[4] An alternative version of the proposition (with a suitably modified Assumption B′) could allow for the parameter values $\frac{1-\sigma}{2} < \mu_A \leq \frac{1}{2}$. In that case, the worker's period-1 maximum effort would be below the efficient level. In the period-1 equilibrium, the firm would offer the lowest wage that elicits that maximal level of effort, and in the period-2 equilibrium, the worker would exert efficient effort and the firm would offer the higher wage that induces it. Therefore, for parameter values in this range, the model predicts that wage and effort will rise over the course of employment. I do not emphasize this prediction because it relies on the worker correctly anticipating the effect of current effort on future fairness judgments, which I believe is much less psychologically plausible than other features of the model.

Part 2 states that the worker's effort does not depend on his period-0 transaction. As in the single-period model, this result follows directly from the worker choosing the efficient level of effort in both periods.

Part 3 of the proposition states that the worker's wage in both periods is increasing in the period-0 material payoff and decreasing in the period-0 profit. The logic hinges on the reference transaction being determined by the previous period's transaction. When the worker's reference payoff is higher and the firm's reference payoff is lower going in to period 1, the firm must pay a higher wage in period 1 to induce the efficient level of effort. Consequently, the worker's period-1 material payoff is higher, and the firm's period-1 profit is lower. Since this means that the worker's reference payoff is higher and the firm's reference payoff is lower going in to period 2, the firm must also pay a higher wage in period 2.[5]

Part 3 of the proposition implies the first motivating fact for this section: workers paid more in period 1 will also tend to be paid more in period 2. Evidence from administrative records indicates that, indeed, cohorts of workers who experience high entry wages continue to earn relatively high wages throughout their tenure at the firm (Baker, Gibbs, and Holmstrom 1994). Beaudry and DiNardo (1991) similarly find that market conditions at the time a worker begins working for a firm has a persistent effect on subsequent earnings (see also Grant 2003, Kahn 2010, and Devereux 2002).

The second motivating fact is that labor market conditions external to the firm do not affect the worker's wage. This is indeed true in the model: the worker's wage path would be unaffected by small variations in the worker's outside-option payoff because as long as the worker is employed, the wage is fully determined by the current output price and the previous period's transaction. The empirical observation that incumbent workers' wages are largely shielded by fluctuations in labor-market supply and demand conditions has been an important theme in the personnel economics literature (Doeringer and Piore 1971, Baker, Gibbs, and Holmstrom 1994, Seltzer and Merrett 2000).[6]

[5] In fact, the model implies that a 1-unit increase in u_0 (or a 1-unit decrease in π_0) has exactly the same effect on w_2^* as it has on w_1^*. I do not emphasize this stronger result because it is sensitive to the simplifying assumption that for a fixed level of total surplus, the fairness function is maximized by *equating* the surpluses. In the more general case of "fairness-kinked preferences" (analyzed in Benjamin 2014), the fairness component of preferences is maximized by making the worker's material payoff an increasing function of profit, but not necessarily the identity function.

[6] Another potential prediction of the model is that salaries of new workers vary with labor market conditions at time of hire. This is implied by the model if it is assumed that the period-0 transaction reflects conditions in the labor market in period 0. Such an assumption is consistent with the evidence presented by Kahneman, Knetsch, and Thaler (1986). Consistent with the prediction, empirical work generally finds that wages of new workers are much more sensitive to labor market conditions than wages of incumbent workers (e.g., Bils 1985, Abowd and Card 1987, Solon, Barsky, and Parker 1994, Baker, Gibbs, and Holmstrom 1994). However, I do not emphasize this prediction because the connection is loose between the assumption that the reference transaction is determined by recent

In labor economics, the two empirical regularities highlighted in this section are often attributed to long-term implicit contracts (e.g., Beaudry and DiNardo 1991). According to the implicit contract interpretation, there is a mutual understanding between the worker and the firm at time of hire about the state-contingent wage path. Labor market conditions at time of hire determine the level of the worker's initial wage, and subsequent labor market conditions are irrelevant because the wage evolves according to the tacitly agreed contract. The implicit contract is often modeled as a reputational equilibrium of a repeated game (e.g., MacLeod and Malcomson 1989). A potentially unsatisfactory aspect of this approach is that such games generally have many equilibria and can flexibly fit a wide variety of possible compensation patterns.

The analysis in this section has shown that worker's fairness concerns can provide an alternative microfoundation for implicit contracts. The empirical regularities arise as the unique equilibrium of the dynamic version of the fairness model. The fairness theory also makes the testable prediction that entry wage persistence and shielding of wages from external labor markets should be observed even in settings where repetition and reputation forces are weak.

5 Loss Aversion and Downward Real Wage Rigidity

While in many countries including the U.S., the predominant pattern of wage stickiness is downward *nominal* wage rigidity, there is also strong evidence for downward *real* wage rigidity, especially in countries with greater union density (Dickens et al 2007). This section explores how workers' concern for fairness—when combined with loss aversion—could provide a plausible account of downward wage rigidity and what additional predictions emerge from such an explanation. Since I defer explicitly modeling the distinction between real and nominal quantities until Section 6, the analysis in this section is best interpreted as relating to downward real wage rigidity.

To keep the analysis as simple as possible, I return to the single-period framework from Section 3, except that (like in Section 4) I assume that p is a random variable drawn from an atomless distribution that has full support on $(1, \infty)$. Moreover, I interpret "period 0" as a time in which the worker was employed at the same firm.

I add to the model "loss aversion," the assumption that losses are weighted more heavily than equivalently-sized gains. Loss aversion is an important feature of preferences in individual decision-making, in both riskless and risky choices (Kahneman and Tversky 1979, Köszegi and Rabin 2006).

personal experience (maintained in the rest of the paper) and the assumption that, when the worker is unemployed, it is determined by conditions in the labor market (needed for the result discussed in this footnote).

While loss aversion has been formalized primarily in models of individual decision-making, available evidence suggests that it also matters for fairness judgments. For example, Kahneman, Knetsch, and Thaler (1986) find that only 20% of respondents consider it unfair for a company to eliminate a ten-percent annual bonus, whereas 61% consider it unfair to reduce wages by ten percent (holding constant total compensation across the two scenarios).[7]

To capture such loss aversion, I assume that the worker weights losses more heavily than gains when calculating his own surplus from the transaction. The evidence for loss aversion in individual decision-making implies that it would also enter into the worker's non-fairness-related utility, but to conserve notation, I incorporate loss aversion only into fairness judgments; in the gift-exchange game I study here, loss aversion in the selfish component of the worker's preferences would affect his willingness to accept employment, but it would not affect his effort choice conditional on employment (and thus would not generate downward wage rigidity) since the worker's effort choice is entirely driven by his fairness concerns.[8]

Formally, I generalize the specification of the worker's surplus payoff as follows: given reference-transaction wage w_0 and effort e_0,

$$\widetilde{u}(w,e) \equiv \Lambda(w - w_0) + \Lambda(-c(e) + c(e_0)),$$

where

$$\Lambda(x) \equiv \begin{cases} x & x \geq 0 \\ \lambda x & x < 0 \end{cases}$$

and $\lambda \geq 1$ is a parameter capturing the degree of loss aversion. The model specializes to the case considered in Section 3 if $\lambda = 1$, but if $\lambda > 1$, the worker weights losses relative to the reference transaction more heavily than gains. I follow Köszegi and Rabin (2006) in assuming that loss aversion matters separately for the two dimensions that affect the worker's material payoff, in this context wage and effort. Thus, the specification implies that a worker dislikes a wage cut more

[7] Given this evidence that reducing a bonus is not perceived as negatively as cutting base pay, it may be puzzling that firms do not pay workers much of their compensation through bonuses. One possible explanation is that bonuses may be (correctly) perceived by workers as less permanent, and thus holding total compensation constant, workers prefer to take a job that offers higher base pay.

[8] In a setting with contractible effort, loss aversion in the selfish component of the worker's preferences could dampen wage adjustments but would still not cause the distribution of wage changes to have a pile-up at zero. At an optimal contract, the firm would set the wage such that, given the firm's preferred level of effort, the worker's participation constraint binds, $U^E = 0$. A change in the output price would cause the firm's preferred level of effort to change, which would require a change in the wage. Even without fairness concerns, if the worker were loss averse over wages, then to keep the worker on the $U^E = 0$ indifference curve, the wage would have to be cut by less when effort falls than it would have to rise when effort increases. In a multi-period model, anticipating the costliness of wage variability over time, the firm would dampen its wage adjustments in response to changes in the output price (as per the logic in Elsby, 2009) —but the firm would nonetheless cut wages at least somewhat in response to any fall in the output price.

than he likes a same-sized raise, and the worker also dislikes increasing his effort more than he likes a same-sized reduction in effort.

Partly to avoid substantially complicating the model but mostly because I suspect it is approximately true, I assume that the worker does not weight losses more than gains when calculating the firm's surplus. That is, the firm's surplus profit is the same as in Section 3: $\widetilde{\pi}(w, e; p) \equiv [pe - p_0 e_0] + [-w + w_0]$.[9] As in Section 3, $f(\widetilde{u}, \widetilde{\pi})$ is given by equation (2).

5.1 The worker's effort choice

As in the analysis in Section 3, given the output price p, there is a maximum level of effort that the worker is willing to exert. But for a wage w below the threshold that induces maximum effort, the worker's optimal effort $e(w, p)$ equates the worker's surplus and the firm's surplus. Thus, as before, effort is increasing in the wage and the output price. Due to loss aversion, however, a wage cut reduces the worker's surplus more than a wage increase raises it. Consequently, effort is more responsive to the wage when the worker is experiencing a wage cut. Similarly, because the worker's surplus is affected more strongly by an increase in effort than by a decrease in effort, effort is more responsive to the wage when the worker is reducing effort.[10]

Lemma 2. *Under Assumption A, if $\lambda > 1$, then for any $p > 1$, there exists a $\overline{w}(p)$ such that for $w_0 < \overline{w}(p)$:*

1. *Effort responds more to wage cuts than to wage increases:* $\dfrac{\lim_{w \uparrow w_0} \frac{\partial e(w,p)}{\partial w}}{\lim_{w \downarrow w_0} \frac{\partial e(w,p)}{\partial w}} > 1$.
2. *Effort is more responsive to wage changes when effort is below the reference level of effort than when effort is below it:* $\dfrac{\lim_{e \uparrow e_0} \frac{\partial e(w,p)}{\partial w}}{\lim_{e \downarrow e_0} \frac{\partial e(w,p)}{\partial w}} > 1$.

If $w_0 \geq \overline{w}(p)$, then $e(w, p)$ is constant in w.

[9] If instead I assumed that the worker's calculation of the firm's surplus *did* have loss aversion over revenue and over the wage payout, then these would not qualitatively affect the prediction of downward wage rigidity and would counteract upward effort rigidity. A fall in the output price would cause the firm to experience a loss in revenue, which would make effort increase by more than it would in the absence of loss aversion over revenue. But cutting the wage would still cause a discontinuous increase in the sensitivity of effort to the wage, making the firm reluctant to cut the wage. Loss aversion over the firm's wage payout would make effort more responsive to wage increases, which would counteract upward effort rigidity. As far as I am aware, there is little evidence on whether people are loss averse over others' surpluses when making fairness judgments involving themselves and others.

[10] It would not be necessary to consider wage increases/decrease separately from effort decreases/increases if effort always changed in the same direction as the wage. In fact, however, Proposition 5 below will show that in this model with loss aversion, effort can change even when the wage does not, and effort may not change when the wage does. Moreover, in the absence of the assumption I make below that the reference transaction is an equilibrium, wage and effort could move in opposite directions.

Part 1 of Lemma 2 shows that the model provides a microfoundation for Akerlof and Yellen's (1990) "fair wage-effort hypothesis," which postulates that effort is more sensitive to the wage when the wage is below a reference wage (which Akerlof and Yellen call the "fair wage") than when the wage is above it.[11] Thus, the large body of evidence discussed by Akerlof and Yellen (1990) is supportive of Part 1. This includes evidence from surveys that managers believe that effort responds more to wage cuts than to raises (e.g., Campbell and Kamlani 1997), from psychology experiments that effort is less responsive to wage increases than to wage decreases (e.g., Walster, Walster, and Berscheid 1977), and from sociological observations of work restrictions in response to wages perceived as too low (e.g., Mathewson 1969). More recently, in economics experiments, Kube, Maréchal, and Puppe (2013) find no effect on effort in response to a wage increase, but they find a decrease in effort in response to a wage cut. I am not aware of any evidence regarding Part 2 of Lemma 2.

5.2 The equilibrium with loss aversion

The basic logic of equilibrium is similar to that from Section 3: the worker chooses effort so as to equate the surpluses, and the firm chooses the wage to maximize the sum of the surpluses. The difference is that the worker's surplus now incorporates loss aversion. Because the sum of the surpluses is maximized, the firm's and worker's marginal rates of substitution (MRSs) between effort and wage *calculated from the surpluses* are equal in equilibrium. The firm's MRS is $\frac{p}{1}$. If the worker is not experiencing a loss in either wage or effort, then the worker's MRS is $\frac{c'(e)}{1}$, and thus the equilibrium effort will satisfy $p = c'(e)$—the same as equating the MRSs calculated from profit and material payoff. If the worker is experiencing a loss in both, then the worker's MRS as calculated from his surplus is $\frac{\lambda c'(e)}{\lambda}$, and thus the equilibrium effort will similarly satisfy $p = c'(e)$. However, if the worker suffers a loss in effort but not wage, then the worker's MRS is $\frac{\lambda c'(e)}{1}$, and the equilibrium effort satisfies $p = \lambda c'(e)$. And if the worker suffers a loss in wage but not effort, then the worker's MRS is $\frac{c'(e)}{\lambda}$, and the equilibrium effort satisfies $p = \frac{c'(e)}{\lambda}$. Note that effort is not efficient in the latter two cases.

In Section 3, the reference transaction was allowed to be arbitrary. In this section, however, my aim is to study *changes* in wage and effort that are due to shocks to the firm's output price. If the previous period's transaction were arbitrary, then wage or effort changes could instead be the result

[11] The "fair wage-effort hypothesis" does not have an analog for Part 2. Note also that the model in this paper differs from Akerlof and Yellen's (1990) framework by specifically identifying the reference wage with its period-0 level.

of non-optimal choices in the previous period. For example, if the period-0 wage were extremely high, then the firm would cut the wage in period 1, regardless of the output price. While such situations may sometimes be of interest, here I impose the restriction on the period-0 transaction that it was an equilibrium of the same game played in period 0.

The reference transaction must therefore fall into one of the cases above: (i) $p_0 = c'(e_0)$, (ii) $p_0 = \lambda c'(e_0)$, or (iii) $p_0 = \frac{c'(e_0)}{\lambda}$. I refer to (i) as the **steady-state case** because in this case, if the output price remained constant ($p = p_0$), then the period-0 equilibrium transaction would also be the period-1 equilibrium transaction. Even though the model is static, I use the language "steady state" because a steady-state equilibrium would be a convergence point in a repeated version of the model absent price changes.

I call (ii) the **recent-increase case** because it corresponds to a situation in which both wage and effort increased in the previous period. It could describe a setting in which the firm is becoming more productive or industry demand is increasing. It is not a "steady state" because if the output price remained constant ($p = p_0$), then the period-0 equilibrium transaction (w_0, e_0) could not be the period-1 equilibrium. If the firm set the same wage $w = w_0$, then the worker's optimal effort choice $e(w, p)$ would equate the period-1 surpluses—but this would differ from the effort choice e_0 that equated the period-0 surpluses because the worker would not be experiencing a loss in period 1.

Analogously, I call (iii) the **recent-decrease case** because it occurs when both wage and effort decreased in the previous period. It would be most frequent when the output price is trending downward (e.g., because demand is declining) or the firm is becoming less productive. Like the recent-increase case, it is not a "steady state."

Proposition 5, the main result of this section, outlines the implications of the model for wage and effort as a function of the price realization and whether the reference transaction is in the steady-state, recent-increase, or recent-decrease case. To sidestep defining an analog of Assumption B (which would be more complex), the proposition focuses on equilibria in which the players transact. To facilitate stating the result, define

$$\left(\underline{p}_{\text{w-rigid}}, \overline{p}_{\text{w-rigid}}\right) \equiv p_0 \cdot \begin{cases} \left(\frac{1}{\lambda^2}, \frac{1}{\lambda}\right) & \text{in the recent-increase case} \\ \left(\frac{1}{\lambda}, 1\right) & \text{in the steady-state case} \\ (1, \lambda) & \text{in the recent-decrease case} \end{cases}$$

and

$$\left(\underline{p}_{\text{e-rigid}}, \overline{p}_{\text{e-rigid}}\right) \equiv p_0 \cdot \begin{cases} \left(\frac{1}{\lambda}, 1\right) & \text{in the recent-increase case} \\ (1, \lambda) & \text{in the steady-state case} \\ \left(\lambda, \lambda^2\right) & \text{in the recent-decrease case} \end{cases}.$$

Proposition 5. *Under Assumption A, if the firm hires the worker in equilibrium, then the equilibrium (w^*, e^*) is unique almost surely. Moreover:*

1. *If $p \in \left(\underline{p}_{\text{w-rigid}}, \overline{p}_{\text{w-rigid}}\right)$, then $w^* = w_0$, $e^* > e_0$, and e^* is strictly decreasing in p.*
2. *If $p \in \left(\underline{p}_{\text{e-rigid}}, \overline{p}_{\text{e-rigid}}\right)$, then $e^* = e_0$, $w^* > w_0$, and w^* is strictly increasing in p.*
3. *If p is outside the above ranges, then w^* and e^* are both strictly increasing in p.*

Part 3 of the proposition implies that for a large enough positive shock, the firm increases the wage and the worker increases effort, and for a large enough negative shock, the firm cuts the wage and the worker reduces effort. The qualitatively distinctive implications of loss aversion arise for relatively small negative or positive shocks, as described in Parts 1 and 2 of the proposition.

Part 1 states that the wage is rigid when the price realization occurs within the interval $\left(\underline{p}_{\text{w-rigid}}, \overline{p}_{\text{w-rigid}}\right)$. Recall from Lemma 2 that effort is more responsive to the wage when $w < w_0$ than when $w \geq w_0$. The value $\overline{p}_{\text{w-rigid}}$ is the price at which the worker chooses effort e_0 when the firm sets the wage w_0. For a price realization just below $\overline{p}_{\text{w-rigid}}$, the profit-maximizing wage is the corner solution w_0. Intuitively, starting from the wage w_0, since the price has fallen, the worker chooses inefficiently high effort ($e(w_0, p) > e_{\text{eff}}(p)$) to keep the surpluses equal. Thus, the firm would like to reduce the wage slightly, but if it did so, then effort would discontinuously become more responsive to the wage, making the optimal wage higher. The value $\underline{p}_{\text{w-rigid}}$ is the price at which the firm is just indifferent between setting the wage w_0 thereby inducing inefficiently high effort and cutting the wage with the consequent sharp reduction in effort. (It is because of this indifference when $p = \underline{p}_{\text{w-rigid}}$ that the proposition states that the equilibrium is unique "almost surely.")

The model makes two novel predictions regarding effort in the region of wage rigidity. First, the level of effort will be *higher* than in the previous period. That is because effort is e_0 when the price realization is exactly $\overline{p}_{\text{w-rigid}}$, and for lower price realizations, with the wage held fixed, effort needs to be higher to keep the surpluses equal. Second, effort is *decreasing* in the output price—the opposite comparative static from when the price shock causes a wage adjustment—because

the lower the price, the higher the equal-surplus effort level. In principle, both of these predictions could provide a way of distinguishing the fairness theory from alternative explanations of downward wage rigidity. In practice, however, it may be difficult to find a clean natural experiment because complications that are omitted from the model may have implications in the opposite direction. For example, in real-world settings there is typically asymmetric information, and workers might reduce effort if they believe management is misrepresenting the firm's profitability. And due to negative reciprocity, if workers perceive management to be responsible for the problems, they may want to punish the firm. Indeed, in one manufacturing company that cut wages studied by Greenberg (1990), apologizing and informing workers that a pay cut was necessitated by financial pressures led to a smaller increase in employee theft. This finding may suggest that blaming the company and not appreciating the severity of the financial problems are forces that cause workers to withdraw effort.

I am aware of a bit of evidence, albeit somewhat indirect, related to the second of these predictions: managers that find it necessary to reduce wages do not seem to suffer consequences as drastic as predicted by other managers who describe what would happen if they cut wages (Bewley 1999). In the equilibrium of the model, if the realized price is just low enough to make it optimal to cut wages, the worker's effort is actually *higher* than if the realized price were the same as the period-0 price.

Another set of distinguishing predictions comes from Part 2 of Proposition 5, which states that the *effort* is rigid when the price realization occurs within the interval $\left(\underline{p}_{\text{e-rigid}}, \overline{p}_{\text{e-rigid}}\right)$. This prediction follows from the result in Lemma 2 that effort is more responsive to the wage when $e \leq e_0$ than when $e > e_0$. The intuition underlying effort rigidity is similar to that for wage rigidity: for any price realization between $\underline{p}_{\text{e-rigid}}$ and $\overline{p}_{\text{e-rigid}}$, starting from whatever wage is needed to induce effort level e_0, effort is inefficiently low ($e_0 < e_{\text{eff}}(p)$) and thus the firm would like to increase the wage, but then effort would become discontinuously less responsive to the wage, making the optimal wage lower. The profit-maximizing wage is a corner solution, the wage level that induces exactly e_0. This wage is higher than in the previous period—even if the price shock is negative—and is increasing in the price. I am not aware of any evidence regarding effort rigidity.

A final set of novel predictions relates to when wage and effort rigidity are predicted to occur. The steady-state case is predicted to exhibit downward wage rigidity in response to a small negative shock and upward effort rigidity in response to a small positive shock. In contrast, in the recent-increase case, there are no rigidities in response to a positive shock because the worker will remain in the domain of a gain in wage ($w \geq w_0$) and a loss in effort ($e > e_0$); thus when the firm increases the

wage, there are no discontinuous changes in the responsiveness of effort. For the price realization $p = p_0$, the firm optimally sets the wage that induces effort e_0, but this wage is higher than w_0 because effort is relatively unresponsive to the wage (since the worker is in the domain of a loss in effort). For a small negative shock, the firm reduces the wage toward w_0, and the equilibrium occurs in the region of effort rigidity. For a somewhat larger negative shock, the equilibrium instead occurs in the region of wage rigidity. The recent-decrease case is analogous to the recent-increase case: there are no rigidities in response to a negative shock, a small positive shock generates wage rigidity, and a somewhat larger positive shock generates effort rigidity. While I am not aware of much evidence regarding this rich set of predictions, I discuss some related evidence in the next section.

6 Money Illusion and Downward Nominal Wage Rigidity

While there is evidence for downward real wage rigidity, the evidence is more widespread for downward nominal wage rigidity (DNWR), which extends across union and non-union firms in a range of countries and inflationary environments (Dickens et al 2007). Figure 1, taken from Fehr and Goette (2005), shows the distribution of wage changes in Switzerland in 1994, a year in which inflation was 1.6%. As is typical, there are more wage increases than decreases, and there is a pile-up of observations at zero wage change. DNWR is a puzzle that may have important consequences for understanding business cycles (Elsby 2004, Collard and de la Croix 2000) and for optimal inflation-targeting (Tobin 1972, Akerlof, Dickens, and Perry 1996).

Fairness plays a prominent role in informal explanations of DNWR.[12] Managers say they avoid wage cuts because workers would perceive them as unfair (Blinder and Choi 1990, Kaufman 1984, Agell and Lundborg 1995) and respond with worse performance (e.g., Bewley 1999, Campbell and Kamlani 1997).

To study DNWR, I make one crucial modification to the model from the previous section: I

[12] A major alternative explanation is that cutting the wage will cause the workers who have the best outside option—the most productive workers—to quit, and it is more profitable to lay off the least productive workers. This explanation makes sense theoretically in environments in which workers with the same job category must be paid equally, which is often true in unionized firms and is sometimes imposed in non-unionized firms by the firm's personnel policies; firms without the equal-pay constraint could instead adjust wages optimally for each individual worker according to his or her productivity. Other alternative explanations seem much less plausible. Menu costs of changing wages (Ball and Mankiw 1994) have difficulty explaining why so many wage changes are small, including wage increases. Moreover, the source of menu costs is not clear in this context. Insurance from an optimal long-term contract against wage reductions (Harris and Holmstrom 1982) has difficulty explaining why wages seem downward rigid in high-turnover jobs and why wage cuts do sometimes occur. Explicit contracts with loss averse workers might prohibit wage reductions (Elsby 2004), but such explicit provisions are unusual. When managers are directly confronted with "fairness" versus other explanations, they typically endorse fairness (e.g., Campbell and Kamlani 1997, Blinder and Choi 1990).

assume that the monetary amounts in the worker's reference transaction are nominal quantities, rather than real quantities. This "money illusion" assumption is consistent with survey evidence on how people make a variety of judgments (Shafir, Diamond, and Tversky 1997), including fairness judgments. For example, survey respondents consider it much less fair for a company to reduce salaries by 7% when inflation is 0% than to increase salaries by only 5% when inflation is 12%, even though the two are equivalent in real terms (Kahneman, Knetsch, and Thaler 1986).

If some workers are not subject to money illusion (perhaps because union leaders explain to workers that they should adjust for inflation) and some workers are, then the analysis from the previous section would apply to the first type of worker, and the analysis from this section would apply to the second type. The evidence that actual wage distributions exhibit both downward real wage rigidity and DNWR would be explained by the population of workers being a mixture of these types.

To formally model fairness concerns with money illusion, I denote the current **price level** by the exogenous constant $P > 0$. The **nominal wage** is Pw, and the **nominal price** of the firm's output is Pp. The reference transaction, $(w_0, e_0; p_0, P_0)$, now includes the period-0 **price level** $P_0 > 0$. I interpret the worker's material payoff as reflecting the utility he gets from consuming the goods he purchases (and his disutility of effort), so it continues to depend only on real quantities: $u(w, e) = w - c(e)$. The worker's fairness function is the same as before, but now the surpluses are nominal. The firm's surplus payoff is the firm's nominal profit relative to its period-0 nominal profit:

$$\widetilde{\pi}(w, e; p, P) \equiv [Ppe - P_0 p_0 e_0] + [-Pw + P_0 w_0].$$

The worker's surplus payoff is

$$\widetilde{u}(w, e; P) \equiv \Lambda(Pw - P_0 w_0) + P\Lambda(-c(e) + c(e_0)).$$

The firm term means that the worker evaluates his gain or loss in the nominal wage. The second term has a different form than the first term because effort is not a monetary amount and is therefore not subject to money illusion: if $e = e_0$, then the second term is zero regardless of how the price level changes. However, the second term needs to be multiplied by the price level because the worker's surplus is measured in nominal units. Because the worker's utility when employed is measured in real units, its fairness component (which is now measured in nominal units) needs to be normalized by the price level: $U^E = \sigma u + (1 - \sigma)\frac{f(\widetilde{u}, \widetilde{\pi})}{P}$. When the price level is constant ($P = P_0$), it cancels out of the worker's utility function, and the model specializes to that from the previous section.

As before, the worker's optimal effort $e(wP, pP)$ equates the worker's surplus and the firm's surplus, but now it depends on the nominal wage and output price. Therefore a nominal analog of Lemma 2 applies: effort is more responsive to nominal wage cuts than nominal wage increases, and it is more responsive to the nominal wage when effort is above its reference level than when it is below.

Assuming that the period-0 transaction was an equilibrium in period 0, the reference transaction can again be put into one of three cases: (i) the steady-state case ($p_0 = c'(e_0)$), (ii), the recent-increase case ($p_0 = \lambda c'(e_0)$), or (iii) the recent-decrease case ($p_0 = \frac{c'(e_0)}{\lambda}$). Now, however, "recent increase," and "recent decrease" refer to what recently happened to effort and the *nominal* wage. Relatedly, the interpretation of "steady state" is different. In order for the period-1 equilibrium transaction (w^*, e^*) to equal the period-0 equilibrium transaction (w_0, e_0), not only the output price needs to remain constant ($p = p_0$) but also the price level needs to remain constant ($P = P_0$).

The model predicts DNWR because Proposition 5 carries over but with the nominal price and nominal wage replacing the real price and real wage everywhere in the proposition. To illustrate the model's predictions about the cross section of wage changes for this case, Figure 2 shows, for a particular specification of the model's parameters, simulation results for how a smooth (log-normal) distribution of price changes translates into the distribution of wage changes.[13] Panels A and B show cases of no loss aversion ($\lambda = 1$) and loss aversion ($\lambda > 1$), respectively. The reference transaction is in the steady-state case, and to parallel the data in Figure 1, the inflation rate is set to 1.6%. While the model is highly stylized, the predicted distribution of wage changes under loss aversion replicates two of the qualitative features of the empirical distribution shown in Figure 1: a spike at zero, and an apparent shift of mass away from the negative-wage-change part of the distribution (into the spike at zero). However, given a population of identical firms and workers, the model *cannot* explain why there are some slightly negative wage changes. As noted after Proposition 5, at the lowest price in the wage-rigidity range, the firm switches from w_0 to a discretely lower wage, and for lower price realizations, the equilibrium wage is even lower. Thus, the model generates a gap in the distribution of wage changes between zero and a slightly negative wage change.[14] With plausible heterogeneity in firms and workers, however, the gap would occur in different places for different firm-worker pairs, and the data averaged over many workers would

[13] The Mathematica code used to generate Figure 2 is available on the author's website.

[14] In a different context (without fairness concerns)—the distribution of U.S. income tax filings—Rees-Jones (2014) has shown that loss aversion generates a similar "shift" and "spike" in the distribution. The underlying mechanics for the prediction are similar to those underlying wage rigidity, but in that context, there is no predicted gap in the distribution of filings.

have no gap.

It is sometimes argued that inflation "greases the wheels" of the labor market by enabling downward real wage adjustments to occur when it is efficient for them to occur. In terms of the model, the nominal analog of Proposition 5 indeed implies that the real wage is downward flexible as long as the nominal wage does not need to be cut. However, the model also implies that inflation causes *redistribution* from workers to firms: because effort responds to the nominal wage, whenever the price level increases, the firm can induce its desired level of effort with a lower real wage.[15] However, since a fall in the real wage reduces the worker's material payoff, this redistribution will ultimately be limited by the firm's need to beat the worker's outside option. The model similarly predicts that deflation causes redistribution from firms to workers because when the price level falls, firms need to pay a higher real wage to induce any given level of effort.

The model also has implications for how the frequency distribution of wage changes depends on the average rate of increase of nominal prices Pp across firms. Starting from an average rate of increase close to zero, a higher rate—corresponding to an economy with a higher rate of inflation or productivity growth, or an industry experiencing growing demand—is predicted to generate a lower frequency of zero wage changes. A lower fraction of firms will be in the recent-decrease case, and hence a greater fraction of firms need to be hit by a negative shock in order to exhibit DNWR. Moreover, fewer firms will be hit by negative shocks. Consequently, fewer firms will be in the region of wage rigidity. Consistent with this prediction, Fehr and Goette's (2005) evidence from Switzerland during 1991-1997 indicates that the frequency of zero nominal wage changes was lower in the early period of substantially higher inflation.

If most firms are experiencing sustained declines in Pp—as would occur in an economic environment of deflation or productivity decline, or an industry with contracting demand—then there will be very few zero wage changes because most firms will be in the recent-decrease case and thus wages will not be rigid when Pp falls. Consistent with this prediction, Kuroda and Yamamoto's (2005) examination of wage changes in Japan suggests that there was DWNR during 1996-1997 at the beginning of the deflation, but wages have been downward flexible since 1998.

[15] The model is therefore a version of what Mankiw (1994, pp. 292-292) calls the "worker-misperception model," originally due to Friedman (1968). This model provides an explanation for why the aggregate supply curve (the relationship between price level and aggregate output) is upward sloping: workers misperceive a rise in the price level as an increase in the real wage and increase their labor supply.

7 Efficiency of the Employment Transaction

Whereas the previous sections of this paper were primarily concerned with positive implications of the fairness model, this section discusses some normative implications. Because the wage is set in order to influence the worker's effort rather than to clear the labor market, the model is an "efficiency wage" model. It therefore raises the same normative issues for the labor market as a whole regarding unemployment as other efficiency-wage models. In this section, I focus attention on normative questions regarding the efficiency of the bilateral transaction between the firm and the worker. To minimize notation, the formal analysis is restricted to the single-period version of the game from Section 5 (i.e., without money illusion).

Following Benjamin (2014), I distinguish between two notions of efficiency for studying interactions when agents have other-regarding preferences. "Material Pareto efficiency" is Pareto efficiency with respect to the purely self-regarding component of preferences (the worker's material payoff and the firm's profit), whereas "utility Pareto efficiency" is Pareto efficiency with respect to the overall objective function that the players maximize. In the present context, a transaction (w, e) is called **material Pareto efficient (MPE)** if there is no alternative transaction (w', e') such that $u(w', e') \geq u(w, e)$ and $\pi(w', e') \geq \pi(w, e)$, at least one inequality strict. A transaction (w, e) is called **utility Pareto efficient (UPE)** if there is no alternative transaction (w', e') such that $U(w', e') \geq U(w, e)$ and $\pi(w', e') \geq \pi(w, e)$, at least one inequality strict. I defer until later in this section a discussion of whether MPE or UPE may be the right social welfare criterion. I focus first on characterizing under what conditions the equilibria described in previous sections are MPE or UPE.

A transaction is MPE if and only if $\frac{\partial u(w,e)/\partial w}{\partial u(w,e)/\partial e} = \frac{\partial \pi(w,e)/\partial w}{\partial \pi(w,e)/\partial e}$. Because the worker's material payoff function and the firm's profit are quasi-linear in the wage, this equality is equivalent to the worker exerting the efficient level of effort, $e_{\text{eff}}(p)$; the wage does not matter for material efficiency because it is merely a transfer between profit and the worker's material payoff. Thus, Proposition 1 immediately implies that in the absence of loss aversion, the equilibrium is MPE.

Proposition 6 shows that, in the absence of loss aversion, the equilibrium is also UPE.

Proposition 6. *Under Assumptions A and B, if the worker is not loss averse ($\lambda = 1$), then for any $p > 1$, the equilibrium transaction is UPE and MPE.*

Proposition 6 follows immediately from Proposition 1 and Benjamin's (2014) Theorem 1. To understand why the equilibrium is UPE, first notice that any UPE transaction must maximize

the sum of surpluses because otherwise it would be utility-Pareto-dominated by an alternative transaction that maximized the sum of surpluses, had a higher wage and higher effort, and were at least as fair. Among the transactions that maximize the sum of surpluses, relative to the equilibrium, the firm would clearly be worse off if the wage were higher, and the worker would be worse off if the wage were lower since his material payoff would be lower *and* the transaction would be less fair.

The conclusion that the equilibrium is both MPE and UPE can be extended straightforwardly to apply also to the equilibrium of the dynamic game described in Proposition 3. Thus, in the absence of loss aversion, gift exchange that is sustained by fairness concerns can be efficient regardless of which efficiency notion is used.

This conclusion no longer holds if the worker is loss averse ($\lambda > 1$). The equilibrium is still UPE *outside the ranges of wage and effort rigidity*—and is MPE only when wage and effort remain unchanged from the previous period.

Proposition 7. *Under Assumption A, if the worker is loss averse ($\lambda > 1$) and the firm hires the worker in equilibrium, then the equilibrium transaction is UPE if and only if* $p \notin \left(\underline{p}_{\text{w-rigid}}, \overline{p}_{\text{w-rigid}}\right) \cup \left(\underline{p}_{\text{e-rigid}}, \overline{p}_{\text{e-rigid}}\right)$. *The equilibrium transaction is also MPE if and only if* $p = \overline{p}_{\text{w-rigid}}$, *or equivalently, if and only if the equilibrium transaction* (w^*, e^*) *satisfies* $w^* = w_0$ *and* $e^* = e_0$.

Outside the ranges of wage and effort rigidity, the logic for why the equilibrium is UPE is similar to the non-loss-averse case. However, if the price realization occurs in $\left(\underline{p}_{\text{w-rigid}}, \overline{p}_{\text{w-rigid}}\right)$, then there is a utility-Pareto improvement: a small reduction in wage and effort that keeps profit constant. Because effort is inefficiently high ($e^* > e_{\text{eff}}(p)$) in the region of wage rigidity, the joint reduction in wage and effort would increase the worker's material payoff and (since profit is held constant) therefore also utility. Similarly, if the price realization occurs in $\left(\underline{p}_{\text{e-rigid}}, \overline{p}_{\text{e-rigid}}\right)$, then there is a utility-Pareto improvement. In this case, because effort is inefficiently *low* ($e^* < e_{\text{eff}}(p)$), a joint, small *increase* in wage and effort that keeps profit constant would increase utility.

Turning to MPE, as noted at the beginning of Section 5.2, the equilibrium level of effort may not be $e_{\text{eff}}(p)$—and thus the equilibrium transaction may not be MPE. In fact, Proposition 5 implies that the equilibrium effort is the efficient level of effort only for the unique price realization at which $w^* = w_0$ and $e^* = e_0$. In the region of wage rigidity, effort is above the efficient level, and in the region of effort rigidity, effort is below it. For any other price realization, either (a) $w^* > w_0$ and $e^* > e_0$, or (b) $w^* < w_0$ and $e^* < e_0$. In case (a), the worker suffers a loss in effort but not wage, and therefore as noted at the beginning of section 5.2, the equilibrium effort satisfies $p = \lambda c'(e^*)$

and hence is below the efficient level. In case (b), the worker suffers a loss in wage but not effort, and therefore the equilibrium effort satisfies $p = \frac{c'(e^*)}{\lambda}$ and is above the efficient level.

Is UPE or MPE the more appropriate generalization of Pareto efficiency to use as a welfare criterion? UPE is the appropriate generalization if the worker's utility represents his "true" preferences, by which I mean what he would choose given correct beliefs and after deliberation. But there are a number of reasons why the worker's behavior—as represented by his utility function—may deviate from his true preferences (for related discussion, see Benjamin 2014, Köszegi and Rabin 2008, and Sen 1973).

In the gift-exchange game studied here, one key question is whether the worker's fair-minded behavior reflects his true preferences, or whether it reflects social norms or reliance on a heuristic (modeled in a reduced-form way via the utility specification) that he would reject upon further deliberation. For example, aiming to share total surplus equally may be a heuristic that he would endorse upon deliberation when interacting with another person but not when interacting with a firm. If the worker's true preference when interacting with a firm coincided with his material payoff, then MPE would be the appropriate generalization of Pareto efficiency.

Even if the worker's fair-minded behavior is a true preference, his loss aversion may be at least partially a mistake. For example, loss aversion might reflect an immediate emotional reaction that would fade with deliberation. Even if the worker, upon deliberation, would choose to avoid current feelings of loss aversion, he might excessively choose to avoid losses under the mistaken belief that his current feelings of loss will persist (Loewenstein, O'Donoghue, and Rabin 2003). If loss aversion in this context is entirely an error, then the appropriate notion of efficiency is UPE but with the worker's utility calculated without loss aversion.[16] (Money illusion is surely an error, even for a worker who truly has loss-averse preferences.)

In my view, theoretical reasoning alone cannot resolve the question of what are the worker's "true" preferences—and indeed the answer may differ from one worker to another. The purpose of the analysis here is to highlight when and how the answer matters. To shed light on which preferences should be used for normative purposes, I believe that empirical evidence must also come into play (for related discussion, see Beshears et al 2008). For example, experimental research

[16] If the worker is loss averse but UPE is defined with respect to the worker's utility calculated without loss aversion, then the equilibrium is UPE if and only if $p = \overline{p}_{\text{w-rigid}}$. To see why, note that Benjamin's (2014) Theorem 1 implies that, with a non-loss-averse utility function, the equilibrium is UPE only if it is MPE. Proposition 7 states that when the worker is loss averse, the equilibrium is MPE only in the special circumstance that $p = \overline{p}_{\text{w-rigid}}$. Moreover, when $p = \overline{p}_{\text{w-rigid}}$, the equilibrium for a loss-averse worker coincides with the equilibrium for a non-loss-averse worker. Proposition 6 states that this equilibrium for a non-loss-averse worker is also UPE with respect to the utility of a non-loss-averse worker.

could examine whether and to what extent information or deliberation weaken loss aversion and fair-minded behavior. The conclusions from such research could be useful far beyond the specific analysis of the firm-worker interaction studied in this paper.

8 Concluding Remarks

The model of fairness concerns explored in this paper—a desire to equalize surplus payoffs judged relative to a reference transaction, which is determined by recent personal experience—also may explain a wide variety of observations about non-labor markets. For example, consumers may wish to punish firms that they believe are trying to extract more surplus from them than usual. Consistent with this hypothesis, during temporary periods of high demand, firms often voluntarily maintain prices below the market-clearing level, leading to long lines or stockouts (e.g., Rotemberg 2005, Olmstead and Rhode 1985, Dacy and Kunreuther 1969). Similarly, when costs increase, firms typically postpone raising prices. If it becomes clear that the cost increase is permanent and firms eventually raise their prices, firms often expend resources to inform customers that their profits have taken a hit. In housing markets, rent increases on new tenants are much more common than rent increases on existing tenants (Genesove 1999, Kahneman, Knetsch, and Thaler 1986).

The model explored in this paper is purposefully kept as simple as possible in order to make transparent the link between the worker's fairness concerns and the model's empirical implications. However, it is also important to examine more realistic models to investigate to what extent the conclusions from the simple model can be generalized.

Throughout the analysis, I have assumed that only workers have a concern for fairness. Results would be similar if firms have an analogous concern for fairness (see Benjamin 2014)—at least if the worker and the firm share the same reference transaction. In fact, however, disagreements about the reference transaction can be important in real-world interactions (indeed, Hart and Moore (2008) argue that the potential for such disagreements is a major motivation for contracting). For example, when there are several reasonable precedents, negotiators seem able to convince themselves that the one most favorable to themselves is the most relevant. This self-serving bias can often cause negotiations to break down (Babcock and Loewenstein 1997). Similar problems could arise if only one party has fairness preferences, but the selfish party does not know what the fair-minded party considers to be the reference transaction.

It would be useful to relax some of the restrictions from Assumption A. As noted in Section 3, if the worker is allowed to have a weaker dislike of advantageous unfairness, then results in the

one-period model would be qualitatively similar, except that the equilibrium transaction would not be MPE. Allowing the worker to have a weaker dislike of disadvantageous unfairness can lead to a qualitatively different outcome, in which the firm exploits the worker's willingness to altruistically provide effort.

Another valuable extension would be to embed the interaction between the firm and the worker in a labor market, in which the players' outside options are determined by the prospects of finding another match. Such a model would make it possible to study how aggregate shocks affect aggregate wages and unemployment when workers' loss aversion generates DNWR. Eliaz and Spiegler (2014) make progress in this direction.

Other important extensions include allowing the firm to be uncertain about the workers' preferences (see footnote 1) and making the worker's effort partially contractible. There has been some progress on both of these fronts in closely related models (e.g., Fehr, Klein, and Schmidt 2007).

References

[1] John M. Abowd and David Card. Intertemporal labor supply and long-term employment contracts. *American Economic Review*, 77(1):50–68, March 1987.

[2] John M. Abowd, Francis Kramarz, and David N. Margolis. High wage workers and high wage firms. *Econometrica*, 67(2):251–333, March 1999.

[3] Jonas Agell and Per Lundborg. Theories of pay and unemployment: Survey evidence from swedish manufacturing firms. *Scandinavian Journal of Economics*, 97:295–307, 1995.

[4] George A. Akerlof, William T. Dickens, and George L. Perry. The macroeconomcis of low inflation. *Brookings Papers on Economic Activity*, (1):1–76, 1996.

[5] George A. Akerlof and Janet L. Yellen. The fair wage-effort hypothesis and unemployment. *Quarterly Journal of Economics*, 105(2):255–283, May 1990.

[6] James Andreoni and B. Douglas Bernheim. Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636, 2009.

[7] Linda Babcock and George Loewenstein. Explaining bargaining impasse: The role of self-serving biases. *Journal of Economic Perspectives*, 11(1):109–126, 1997.

[8] George Baker, Michael Gibbs, and Bengt Holmstrom. The internal economics of the firm: Evidence from personnel data. *Quarterly Journal of Economics*, 109(4):881–919, November 1994.

[9] George Baker, Michael Gibbs, and Bengt Holmstrom. The wage policy of a firm. *Quarterly Journal of Economics*, 109(4):921–955, November 1994.

[10] Laurence Ball and N. Gregory Mankiw. Asymmetric price adjustment and economic fluctuations. *Economic Journal*, 104(423):247–261, March 1994.

[11] Paul Beaudry and John DiNardo. The effect of implicit contracts on the movement of wages over the business cycle: Evidence from micro data. *Journal of Political Economy*, 99(4):665–688, August 1991.

[12] Daniel J. Benjamin. A theory of fairness in labor markets. *Harvard University mimeo*, November 2005.

[13] Daniel J. Benjamin. Distributional preferences, reciprocity-like behavior, and efficiency in bilateral exchange. *American Economic Journal: Microeconomics*, Forthcoming.

[14] John Beshears, James J. Choi, David Laibson, and Brigitte C. Madrian. How are preferences revealed? *Journal of Public Economics*, 92(8-9):1787–1794, 2005.

[15] Truman F. Bewley. *Why Wages Don't Fall During a Recession.* Harvard University Press, Cambridge, MA, 1999.

[16] Mark J. Bils. Real wages over the business cycle: Evidence from panel data. *Journal of Political Economy*, 93:666–689, 1985.

[17] Alan S. Blinder and Don H. Choi. A shred of evidence on theories of wage stickiness. *Quarterly Journal of Economics*, 105(4):1003–1015, November 1990.

[18] Lisa E. Bolton, Luk Warlop, and Joseph W. Alba. Explorations in price (un)fairness. *Journal of Consumer Research*, 29:474–91, March 2003.

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

[20] Carl M. Campbell III and Kunal S. Kamlani. The reasons for wage rigidity: Evidence from a survey of firms. *Quarterly Journal of Economics*, 112(3):759–789, August 1997.

[21] David Card, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6):2981–3003, 2012.

[22] Gary Charness and Peter Kuhn. Does pay inequality affect worker effort? experimental evidence. *Journal of Labor Economics*, 25(4):693–723, 2007.

[23] Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869, August 2002.

[24] Alain Cohn, Ernst Fehr, and Lorenz Goette. Fair wages and effort provision: Combining evidence from the lab and the field. 2013. University of Zurich Department of Economics Working Paper Series No. 107.

[25] Fabrice Collard and David de la Croix. Gift exchange and the business cycle: The fair wage strikes back. *Review of Economic Dynamics*, 3:166–193, 2000.

[26] D.C. Dacy and H. Kunreuther. *The economics of natural disasters.* Free Press, New York, 1969.

[27] Paul J. Devereux. The importance of obtaining a high paying job. 2002. UCLA Working Paper.

[28] William T. Dickens, Lorenz Goette, Erica L. Groshen, Steinar Holden, Julian Messina, Mark E. Schweitzer, Jarkko Turunen, and Melanie E. Ward. How wages change: Micro evidence from the international wage flexibility project. *Journal of Economic Perspectives*, 21(2):195–214, 2007.

[29] William T. Dickens and Lawrence F. Katz. Inter-industry wage differences and theories of wage determination. *NBER Working Paper No. 2271*, June 1987.

[30] Peter B. Doeringer and Michael J. Piore. *Internal Labor Markets and Manpower Analysis.* D.C. Heath and Company, Lexington, MA, 1971.

[31] Ronald G. Ehrenberg and George T. Milkovich. Compensation and firm performance. *NBER Working Paper No. 2145*, February 1987.

[32] Kfir Eliaz and Ran Spiegler. Reference dependence and labor market fluctuations. In Jonathan A. Parker and Michael Woodford, editors, *NBER Macroeconomics Annual 2013*, pages 159–200. University of Chicago Press, 2014.

[33] Michael W. L. Elsby. Evaluating the economic significance of downward nominal wage rigidity. August 2004. Working paper.

[34] Michael W. L. Elsby. Evaluating the economic significance of downward nominal wage rigidity. *Journal of Monetary Economics*, 56(2):154–169, 2009.

[35] Constanca Esteves-Sorenson and Rosario Macera. Revisiting gift exchange: Theoretical considerations and a field test. 2013. Yale University Working Paper.

[36] Constanca Esteves-Sorenson, Rosario Macera, and Robert Broce. Do monetary incentives crowd out intrinsic motivation? a field test in the workplace. August 2014. Yale University Working Paper.

[37] Ernst Fehr and Armin Falk. Wage rigidity in a competitive incomplete contract market. *Journal of Political Economy*, 107:106–134, 1999.

[38] Ernst Fehr and Lorenz Goette. Robustness and real consequences of nominal wage rigidity. *Journal of Monetary Economics*, 52:779–804, 2005.

[39] Ernst Fehr, Lorenz Goette, and Christian Zehnder. A behavioral account of the labor market: The role of fairness concerns. *Annual Review of Economics*, 1:355–84, 2009.

[40] Ernst Fehr, Georg Kirchsteiger, and Arno Riedl. Does fairness prevent market clearing? an experimental investigation. *Quarterly Journal of Economics*, 108(2):437–459, 1993.

[41] Ernst Fehr, Georg Kirchsteiger, and Arno Riedl. Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, 42:1–34, 1998.

[42] Ernst Fehr, Alexander Klein, and Klaus M. Schmidt. Fairness and contract design. *Econometrica*, 75(1):121–154, 2007.

[43] Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868, August 1999.

[44] Milton Friedman. The role of monetary policy. *American Economic Review*, 68:1–17, March 1968.

[45] David Genesove. The nominal rigidity of apartment rents. *NBER Working Paper No. 7137*, May 1999.

[46] Uri Gneezy and John List. Putting behavioral economics to work: Field evidence on gift exchange. *Econometrica*, 74(5):1365–1384, 2006.

[47] Darren Grant. The effect of implicit contracts on the movement of wages over the business cycle: Evidence from the national longitudinal surveys. *Industrial and Labor Relations Review*, 56(3):393–408, April 2003.

[48] Jerald Greenberg. Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, 75(5):561–568, 1990.

[49] N. Gregory Gregory. *Macroeconomics*. Worth Publishers, New York, NY, 1994.

[50] Milton Harris and Bengt Holmstrom. A theory of wage dynamics. *Review of Economic Studies*, 49(3):315–333, July 1982.

[51] Oliver Hart and John Moore. Contracts as reference points. *Quarterly Journal of Economics*, 123(1):1–48, February 2008.

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

[53] Holger Herz and Dmitry Taubinsky. Market experience is a reference point in judgments of fairness. 2013. University of Zurich, Department of Economics Working Paper No. 128.

[54] Lisa Kahn. The long-term labor market consequences of graduating from college in a bad economy. *Labour Economics*, 17(2):303–316, 2010.

[55] Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler. Fairness as a constraint on profit seeking entitlements in the market. *American Economic Review*, 76(4):728–41, 1986.

[56] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–91, 1979.

[57] Roger T. Kaufman. On wage stickiness in britain's competitive sector. *British Journal of Industrial Relations*, 22:101–12, 1984.

[58] Botond Koszegi and Matthew Rabin. A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4):1133–1166, 2006.

[59] Botond Koszegi and Matthew Rabin. Choices, situations, and happiness. *Journal of Public Economics*, 92:1821–1832, 2008.

[60] David Kruse. Does profit sharing affect productivity? *NBER Working Paper No. 4542*, November 1993.

[61] Douglas Kruse, Richard Freeman, Joseph Blasi, Robert Buchele, Adrin Scharf, Loren Rodgers, and Chris Mackin. Motivating employee-owners in esop firms: Human resource policies and company performance. *NBER Working Paper No. 10177*, December 2003.

[62] Sebastian Kube, Clemens Puppe, and Michel Mar'e9chal. Do wage cuts damage work morale? evidence from a natural field experiment. *Journal of the European Economic Association*, 11:853–870, 2013.

[63] Sachiko Kuroda and Isamu Yamamoto. Wage fluctuations in japan after the bursting of the bubble economy: Downward nominal wage rigidity, payroll, and the unemployment rate. *Monetary and Economic Studies*, 23(2):1–30, 2005.

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

[65] David I. Levine. Fairness, markets, and ability to pay: Evidence from compensation executives. *American Economic Review*, 83(5):1241–1259, December 1993.

[66] George Loewenstein, Ted O'Donoghue, and Matthew Rabin. Projection bias in predicting future utility. *Quarterly Journal of Economics*, 118(4):1209–1248, 2003.

[67] B. MacLeod and J.M. Malcolmson. Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica*, 57:312–22, 1989.

[68] Alexandre Mas. Pay, reference points, and police performance. *Quarterly Journal of Economics*, 121, 2006.

[69] Stanley B. Mathewson. *Restriction of Output Among Unorganized Workers*. Southern Illinois University Press, Carbondale, IL, 2nd edition, 1969.

[70] Sandra Maximiano, Randolph Sloof, and Joep Sonnemans. Gift exchange in a multi-worker firm. *Economic Journal*, 117:1025–1050, 2007.

[71] A.L. Olmstead and P. Rhode. Rationing without government: The west coast gas famine of 1920. *American Economic Review*, 75:1044–55, 1985.

[72] Canice Prendergast. The provision of incentives in firms. *Journal of Economic Literature*, 37(1):7–63, March 1999.

[73] Robert D. Pritchard, Marvin D. Dunnette, and Dale O. Jorgenson. Effects of perceptions of equity and inequity on worker performance and satisfaction. *Journal of Applied Psychology*, 56(1):75–94, 1972.

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

[75] Alex Rees-Jones. Loss aversion motivates tax sheltering: Evidence from u.s. tax returns. December 2014. Wharton Working Paper.

[76] Julio J. Rotemberg. Fair pricing. March 2005. Harvard Business School mimeo.

[77] Andrew Seltzer and David Merrett. Personnel policies at the union bank of australia: Evidence from the 1888-1900 entry cohorts. *Journal of Labor Economics*, 18(4):573–613, October 2000.

[78] Amartya Sen. Behaviour and the concept of preference. *Economica*, 40(159):241–259, 1973.

[79] Eldar Shafir, Peter Diamond, and Amos Tversky. Money illusion. *Quarterly Journal of Economics*, 112(2):341–74, 1997.

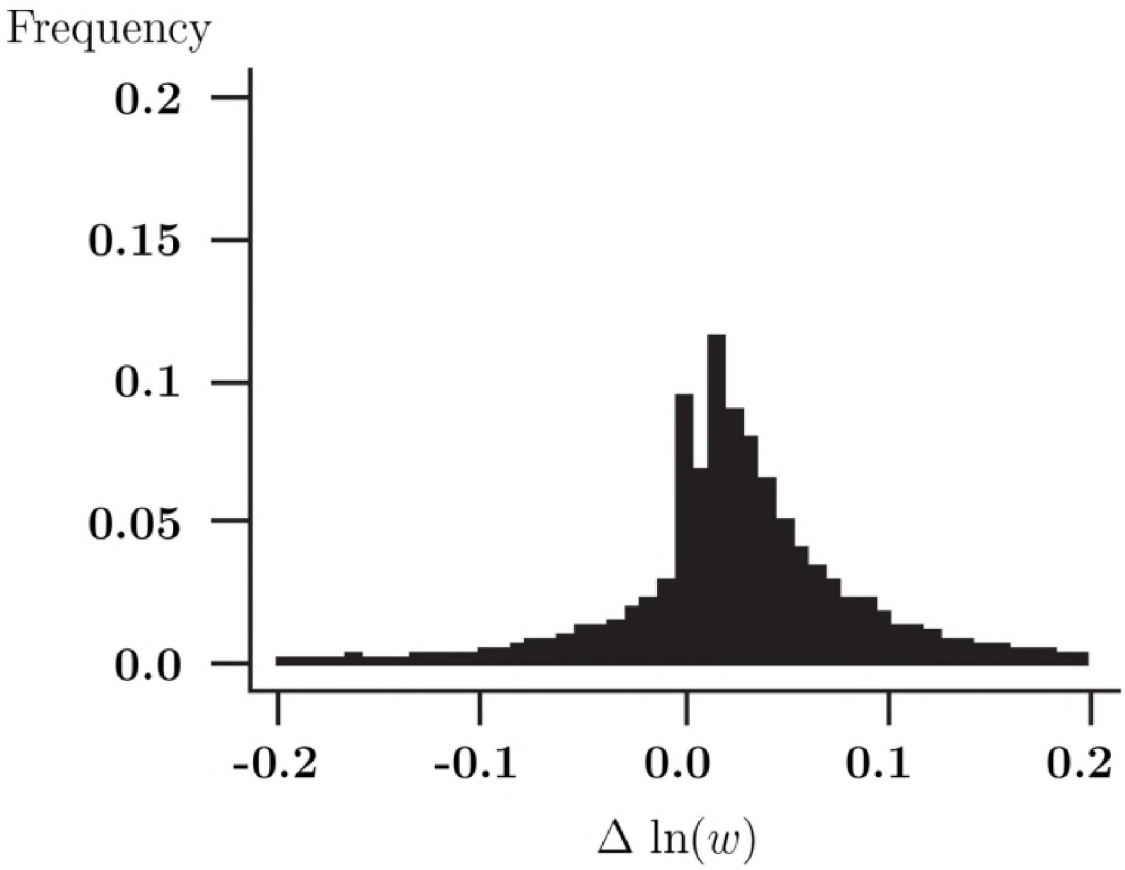
[80] Gary Solon, Robert Barsky, and Jonathan A. Parker. Measuring the cyclicality of real wages: How important is composition bias? *Quarterly Journal of Economics*, 109(1):1–25, 1994.

[81] Richard Thaler. Mental accounting and consumer choice. *Marketing Science*, 4(3):199–214, 1985.

[82] James Tobin. Inflation and unemployment. *American Economic Review*, 62(1/2):1–18, March 1972.

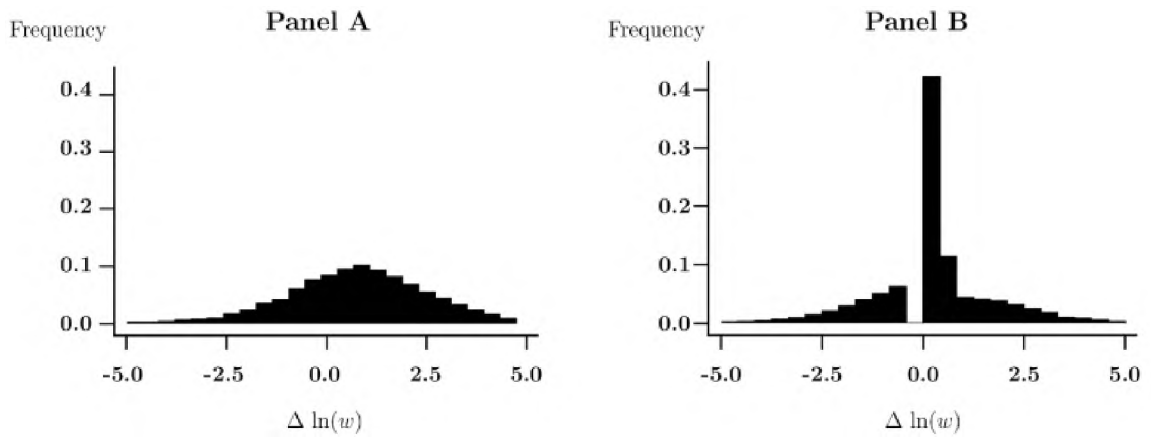
[83] Elaine Walster, G. William Walster, and Ellen Berscheid. *Equity: Theory and Research.* Allyn and Bacon, Boston, 1977.

[84] Martin L. Weitzman and Douglas Kruse. Profit sharing and productivity. In Alan Blinder, editor, *Paying for Productivity.* Brookings Institution, Washington, D.C., 1990.



Source: Fehr and Goette (2005, Fig. 3), as calculated from the Swiss Social Insurance Files.

Figure 1. The distribution of log-wage changes in Switzerland during 1994, a year when the CPI inflation rate was 1.6%. The width of each bin is 0.0083.



Source: Author's calculations.

Figure 2. A simulated distribution of log-wage changes from the model for cases of no loss aversion (Panel A: $\lambda = 1$) and loss aversion (Panel B: $\lambda = 2$, a typical value in the literature) with the reference transaction in the steady-state case. Inflation is set equal to 1.6%: $P/P_0 = 1.016$. Real prices, wages, and the price level at the beginning of the year are normalized to 1: $p_0 = w_0 = P_0 = 1$. Model parameters are: $\ln(p) \sim N(0,1)$, $c(e) = e^2/4$, $\sigma = 0.8$, $\mu_D = 1$, and $\mu_A = 0.5$. The firm and worker are assumed to always transact, regardless of the price realization, as long as the wage is positive. The number of simulation runs (including those dropped due to a negative wage) is 10,000. The width of each bin is 0.43.

Appendix: Proofs

In what follows, there are several useful, alternative ways of expressing the worker's utility when employed, given here for reference (for the model with no loss aversion and no money illusion):

$$\begin{aligned} U^E(w,e) &= \sigma u(w,e) + (1-\sigma) f(\tilde{u}(w,e), \tilde{\pi}(w,e;p)) \\ &= \mu\tilde{u} + (1-\mu)\tilde{\pi} + \sigma u_0 \\ &= (1-\mu)pe - \mu c(e) + (2\mu - 1)w - (\mu - \sigma)u_0 - (1-\mu)\pi_0, \end{aligned}$$

where $\mu \in \{\mu_D, \mu_A\}$.

Lemma 1. *Under Assumption A, for any $p > 1$, there exists $\overline{w}(p)$ such that:*

1. *If $w \leq \overline{w}(p)$, then $e(w,p) \equiv \tilde{e}(p,w)$. Moreover, $\tilde{e}(p,w)$ is increasing in w and decreasing in p.*
2. *If $w > \overline{w}(p)$, then $e(w,p) \in [e_{\text{eff}}(p), \tilde{e}(w,p))$, and $e(w,p)$ is constant in w and increasing in p.*

Proof: Note that $\tilde{\pi}(w,e;p) \leq \tilde{u}(w,e)$ is equivalent to

$$c(e) + pe \leq 2w + \pi_0 - u_0. \tag{1}$$

Furthermore, the left-hand side is strictly increasing in e since $c' > 0$ and $p > 1$. Therefore, $\tilde{\pi}(w,e;p) \leq \tilde{u}(w,e)$ is equivalent to $e \leq \tilde{e}$.

The worker's effort level solves $e(w,p) = \arg\max_e U^E(w,e;p)$. Since U^E is kinked at $\tilde{u} = \tilde{\pi}$, which corresponds to $e = \tilde{e}$, we have

$$\max_e U^E(w,e;p) \equiv \max\Big\{ \underbrace{\max_{e<\tilde{e}} (1-\mu_A)pe - \mu_A c(e)}_{\text{(I)}}, \underbrace{\max_{e\geq\tilde{e}} (1-\mu_D)pe - \mu_D c(e)}_{\text{(II)}} \Big\}.$$

First consider sub-problem (II), and define e_{II} to be the solution to the sub-problem. Assumption A(i) states $\mu_D \geq 1$. Then, on $e \geq \tilde{e}$,

$$\frac{\partial U^E}{\partial e} = (1-\mu_D)p - \mu_D c'(e) < 0.$$

That is, since $\mu_D \geq 1$, U^E is strictly decreasing in e for all $e \geq \tilde{e}$. Hence the maximum is achieved at $e_{\text{II}} = \tilde{e}$.

Consider maximization problem (I) and denote by e_{I} the solution to this sub-problem. In particular, $e_{\text{I}} > 0$ is given by the first-order condition

$$c'(e_{\text{I}}) = p\frac{1-\mu_A}{\mu_A} \geq p \tag{2}$$

since A(ii) implies $\frac{1-\mu_A}{\mu_A} \geq 1$. (Since $\frac{\partial^2 U^E}{\partial e^2} = -\mu_A c''(e)$, $c'' > 0$, and $\mu_A > 0$, the second-order condition for maximization is satisfied on $e > 0$.)

Define $\overline{w}(p)$ by

$$\tilde{e}(p, \overline{w}(p)) = e_{\text{I}}. \tag{3}$$

Since $\tilde{e}$ is increasing in w (see below), for all $w \leq \overline{w}(p)$ we have $e_{\text{I}} \geq \tilde{e}$, and so (I) does not have a solution (because the objective function is strictly increasing in e for all $e < \tilde{e}$). Therefore, if $w \leq \overline{w}(p)$, then the worker's effort level is the solution from (II): $e(w,p) = \tilde{e}$. If $w > \overline{w}(p)$, then (I) has solution e_{I}, which gives higher utility than the solution to (II), $\tilde{e}$, since

$$(1-\mu_A)pe_{\text{I}} - \mu_A c(e_{\text{I}}) > (1-\mu_A)p\tilde{e} - \mu_A c(\tilde{e}) > (1-\mu_D)p\tilde{e} - \mu_D c(\tilde{e}),$$

where the first inequality follows from e_{I} being the solution to (I), and the second inequality follows from $\mu_D > \mu_A$. Therefore, $e(p,w) = e_{\text{I}}$.

Summarizing,

$$e(w,p) = \begin{cases} \tilde{e}(w,p) & \text{if } w \leq \overline{w}(p) \\ e_{\text{I}}(p) & \text{if } w > \overline{w}(p). \end{cases}$$

Equation (2) shows that e_{I} is constant in w and increasing in p. By the definition of $\overline{w}(p)$, $e_{\text{I}}(p) = \tilde{e}(w,p)$ when $w = \overline{w}(p)$, and since $\tilde{e}$ is increasing in w, $e_{\text{I}} < \tilde{e}$ when $w > \overline{w}(p)$. Note that equation (2), A(ii), and $w > \overline{w}(p)$ imply that $e_{\text{I}} \geq e_{\text{eff}}(p)$. Putting these inequalities together: when $w > \overline{w}(p)$, $e_{\text{I}}(p) \in [e_{\text{eff}}(p), \tilde{e}(w,p))$.

Finally, analysis of equation (2) when it holds with equality shows that if we increase w, we must also increase e to achieve equality; i.e., $\frac{\partial \tilde{e}(w,p)}{\partial w} > 0$. Similarly, increasing p increases the left-hand side, so e must be decreased to keep equality; i.e., $\frac{\partial \tilde{e}(w,p)}{\partial p} < 0$.

□

Proposition 1. *Under Assumptions A and B, for any $p > 1$, there is a unique equilibrium in which the firm hires the worker, and the equilibrium transaction (w^*, e^*) satisfies $\pi(w^*, e^*; p) - \pi_0 = u(w^*, e^*) - u_0$ and $e^* = e_{\text{eff}}(p)$.*

Proof: Recall from Lemma 1 that

$$e(w,p) = \begin{cases} \tilde{e}(w,p) & w \leq \overline{w}(p) \\ e_{\mathrm{I}}(p) & w > \overline{w}(p). \end{cases}$$

Therefore, the firm's maximization problem is

$$\max_{w} \pi(w, e(w,p); p) = \max \Big\{ \underbrace{\max_{w \leq \overline{w}(p)} p\tilde{e}(w,p) - w}_{\text{(A)}}, \underbrace{\max_{w > \overline{w}(p)} pe_{\mathrm{I}}(p) - w}_{\text{(B)}}, 0 \Big\}$$

where (A) and (B) are subject to the employment constraint, $U^E(w,e) \geq 0$. The 0 corresponds to the firm's outside option. Let w_{A} and w_{B} denote the solutions to (A) and (B), respectively.

For sub-problem (A), the first-order condition implies

$$\frac{\partial \tilde{e}(w_{\mathrm{A}}, p)}{\partial w} = \frac{1}{p}.$$

Moreover, implicitly differentiating equation (2) with equality with respect to w, we have

$$c'(\tilde{e}(w,p)) \frac{\partial \tilde{e}(w,p)}{\partial w} + p \frac{\partial \tilde{e}(w,p)}{\partial w} = 2.$$

Combining, we have $c'(\tilde{e}(w_{\mathrm{A}}, p)) = p$. That is, $\tilde{e}(w_{\mathrm{A}}, p) = e_{\mathrm{eff}}(p)$. To see that the second-order condition for a global maximum is satisfied, note that an increase in w increases $\tilde{e}$, so $c'(\tilde{e})$ also increases. To maintain the equality, it must be that $\frac{\partial \tilde{e}}{\partial w}$ decreases, hence $\frac{\partial^2 \tilde{e}}{\partial w^2} < 0$. It follows that $\frac{\partial^2 \pi}{\partial w^2} = p \frac{\partial^2 \tilde{e}}{\partial w^2} < 0$. Therefore, the unique solution to (A) is given by $e_{\mathrm{A}} = \tilde{e}(w_{\mathrm{A}}, p) = e_{\mathrm{eff}}(p)$ and w_{A} such that $\tilde{u}(w_{\mathrm{A}}, e_{\mathrm{A}}) = \tilde{\pi}(w_{\mathrm{A}}, e_{\mathrm{A}}; p)$, which implies

$$w_{\mathrm{A}} = \frac{c(e_{\mathrm{eff}}(p)) + pe_{\mathrm{eff}}(p) - \pi_0 + u_0}{2}. \tag{4}$$

We now check that both players choose to interact rather than taking their outside options. The worker chooses employment whenever $U^E(w_{\mathrm{A}}, e_{\mathrm{eff}}(p)) \geq 0$. Since

$$\begin{aligned} U^E(w_{\mathrm{A}}, e_{\mathrm{eff}}(p)) &= \sigma u(w_{\mathrm{A}}, e_{\mathrm{eff}}) + (1-\sigma)\left(u(w_{\mathrm{A}}, e_{\mathrm{eff}}) - u_0\right) \\ &= w_{\mathrm{A}} - c(e_{\mathrm{eff}}) - (1-\sigma)u_0 \\ &= \frac{c(e_{\mathrm{eff}}(p)) + pe_{\mathrm{eff}}(p) - \pi_0 + u_0}{2} - c(e_{\mathrm{eff}}) - (1-\sigma)u_0 \\ &= \frac{M(p)}{2} - \frac{\pi_0}{2} + \frac{2\sigma - 1}{2} u_0, \end{aligned}$$

accepting the offer is equivalent to $(2\sigma - 1)u_0 - \pi_0 \geq -M(p)$. On the other hand, the firm offers the wage only if $\pi(w_{\mathrm{A}}, e_{\mathrm{eff}}(p); p) \geq 0$. Since

$$\pi(w_{\mathrm{A}}, e_{\mathrm{eff}}(p)) = pe_{\mathrm{eff}}(p) - w_{\mathrm{A}} = \frac{pe_{\mathrm{eff}}(p) - c(e_{\mathrm{eff}}(p)) + \pi_0 - u_0}{2},$$

making the offer is equivalent to $u_0 - \pi_0 \leq M(p)$. Both conditions are implied by Assumption B.

We now check that the candidate solution is in fact an interior solution. Note that $w_{\mathrm{A}} \leq \overline{w}(p)$ is equivalent to $e_{\mathrm{I}} \geq \tilde{e}(w_{\mathrm{A}}, p) = e_{\mathrm{eff}}(p)$, which is guaranteed by A(ii) and $c'' > 0$ since e_{I} is defined by equation (2). Therefore (A) has an interior solution.

Clearly if (B) has no solution, then the solution to (A) is the solution to the maximization problem. So suppose (B) has a solution $w_B > \overline{w}(p)$. (Note that even though the maximand in (B) is strictly decreasing, a solution defined by the employment constraint $U^E = 0$ may exist.) We show that in this case, the solution to (A) dominates the solution to (B). Since w_A is the solution to (A), it must necessarily be the case that

$$p\tilde{e}(w_{\mathrm{A}}, p) - w_{\mathrm{A}} \geq p\tilde{e}(\overline{w}(p), p) - \overline{w}(p).$$

Since $\overline{w}(p)$ is defined by $\tilde{e}(\overline{w}(p), p) = e_{\mathrm{I}}$ in equation (3) and $w_B > \overline{w}(p)$, we have

$$p\tilde{e}(\overline{w}(p), p) - \overline{w}(p) = pe_{\mathrm{I}} - \overline{w}(p) > pe_{\mathrm{I}} - w_{\mathrm{B}},$$

Combining, we see that the solution to (A) dominates the solution to (B) whenever the latter exists.

□

Proposition 2. *At the equilibrium described in Proposition 1:*

1. w^* *and* e^* *are both increasing in* p.
2. e^* *does not depend on* u_0 *nor* π_0.
3. w^* *is increasing in* u_0 *and decreasing in* π_0.

Proof: From the proof of Proposition 1, we know that w^* is given by equation (4) and $e^* = e_{\mathrm{eff}}(p)$. Since $e_{\mathrm{eff}}(p)$ is defined by $c'(e_{\mathrm{eff}}(p)) = p$ and since $c'' > 0$, $e_{\mathrm{eff}}(p)$ is increasing in p and is constant in both u_0 and π_0. Given equation (4), w^* is clearly decreasing in π_0 and increasing in u_0. Furthermore, since $e_{\mathrm{eff}}(p)$ is increasing in p and since $c' > 0$, w^* is also increasing in p.

□

Proposition 3. *Under Assumptions* A' *and* B'*, with positive probability there is a unique equilibrium in which the firm hires the worker in both periods. The equilibrium transactions,* (w_t^*, e_t^*) *for* $t = 1, 2$*, satisfy* $\pi(w_t^*, e_t^*; p_t) - \pi_{t-1} = u(w_t^*, e_t^*) - u_{t-1}$ *and* $e_t = e_{\mathrm{eff}}(p_t)$.

Proof: We proceed by backwards induction.

First, note that Assumption A is implied by Assumption A′, and there is positive probability that the realized value of p_2 satisfies Assumption B. In that case, the equilibrium in period 2 is given by Proposition 1, where the reference transactions, u_1 and π_1, are the period-1 outcomes. Since $M(p)$ is strictly increasing in p, we can define p^* as the smallest value of p that satisfies Assumption B for u_1 and π_1. Assuming from now on that $p_2 \geq p^*$, $e_2^* = e_{\text{eff}}(p_2)$ and $\pi(w_2^*, e_2^*; p_2) - \pi_1 = u(w_2^*, e_2^*) - u_1$, which imply that

$$w_2^* = \frac{c(e_{\text{eff}}(p_2)) + p_2 e_{\text{eff}}(p_2) - \pi_1 + u_1}{2}.$$

The resulting profit and utility are

$$\begin{aligned}
\pi_2 &= p_2 e_2^* - w_2^* \\
&= p_2 e_2^* - \frac{c(e_{\text{eff}}(p_2)) + p_2 e_{\text{eff}}(p_2) - \pi_1 + u_1}{2} \\
&= \frac{M(p_2) + \pi_1 - u_1}{2}
\end{aligned}$$

and

$$\begin{aligned}
U_2^E &= \sigma u(w_2^*, e_2^*) + (1-\sigma) f(\tilde{u}(w_2^*, e_2^*), \tilde{\pi}(w_2^*, e_2^*)) \\
&= \sigma u(w_2^*, e_2^*) + (1-\sigma)\tilde{u}(w_2^*, e_2^*) \\
&= u(w_2^*, e_2^*) - (1-\sigma)u_1 \\
&= w_2^* - c(e_2^*) - (1-\sigma)u_1 \\
&= \frac{c(e_2^*) + p e_2^* - \pi_1 + u_1}{2} - c(e_2^*) - (1-\sigma)u_1 \\
&= \frac{M(p_2) - \pi_1 + (2\sigma - 1)u_1}{2}.
\end{aligned}$$

Turning our attention to period 1, we first consider the worker's optimization problem. For $\mu \in \{\mu_A, \mu_D\}$, the worker's objective function is given by

$$\begin{aligned}
U_1^E + EU_2^E &= \mu\tilde{u}_1 + (1-\mu)\tilde{\pi}_1 + \sigma u_0 + \frac{1}{2}EM(p_2) - \frac{1}{2}\pi_1 + \frac{2\sigma - 1}{2}u_1 \\
&= \frac{2\mu + 2\sigma - 1}{2}u_1 + \frac{1 - 2\mu}{2}\pi_1 - (\mu - \sigma)u_0 - (1-\mu)\pi_0 + \frac{1}{2}EM(p_2).
\end{aligned}$$

Rescaling, and noting that w_1, u_0, π_0, and $EM(p_2)$ are constants (with respect to the maximiza-

tion), the worker's maximization problem takes the form

$$\max_e U_1^E(w,e) + EU_2^E \equiv \max\Bigg\{ \underbrace{\max_{e<\tilde{e}}(1-2\mu_A)p_1 e - (2\mu_A + 2\sigma - 1)c(e)}_{\text{(I)}}, \underbrace{\max_{e\geq\tilde{e}}(1-2\mu_D)p_1 e - (2\mu_D + 2\sigma - 1)c(e)}_{\text{(II)}} \Bigg\}.$$

Consider sub-problem (II) first. Note that A′(i) implies $1-2\mu_D < 0$ and $2\mu_D + 2\sigma - 1 > 0$. Therefore, the worker's objective function is strictly decreasing on $e \geq \tilde{e}$. Thus, the solution to (II) is $e_{\text{II}} = \tilde{e}$.

For sub-problem (I), A′(ii) implies $1 - 2\mu_A \geq 0$. If $2\mu_A + 2\sigma - 1 \leq 0$ (i.e., $\mu_A \leq \frac{1-2\sigma}{2} < \frac{1}{2}$ since $\sigma > 0$), then the maximand of (I) is strictly increasing in e. In that case, (I) has no solution. On the other hand, if $2\mu_A + 2\sigma - 1 > 0$ (i.e., $\mu_A > \frac{1-2\sigma}{2}$), then the worker's first-order condition implies

$$c'(e_{\text{I}}) = \frac{1-2\mu_A}{2\mu_A + 2\sigma - 1} p_1. \tag{5}$$

(The second-order condition is satisfied in this case since $2\mu_A + 2\sigma - 1 > 0$ and $c'' > 0$.) We conclude that the solution to the worker's problem in period 1 is

$$e_1(w,p_1) = \begin{cases} e_{\text{I}} & \text{if } \mu_A > \frac{1-2\sigma}{2} \text{ and } w > \overline{w}(p_1) \\ \tilde{e} & \text{otherwise,} \end{cases}$$

where (as in the proof of Lemma 1) $\overline{w}(p)$ is defined by $\tilde{e}(p, \overline{w}(p)) = e_{\text{I}}$.

We now turn to the firm's period-1 maximization problem. The firm's profit is given by

$$\begin{aligned} \pi_1 + E\pi_2 &= \pi_1 + \frac{1}{2}EM(p_2) + \frac{1}{2}\pi_1 - \frac{1}{2}u_1 \\ &= \frac{3}{2}p_1 e_1 - 2w_1 + \frac{1}{2}c(e_1) + \frac{1}{2}EM(p_2). \end{aligned}$$

Again, since $EM(p_2)$ is a constant, we will drop this term from the maximization problems in the following.

Case 1. $\mu_A \leq \frac{1-2\sigma}{2}$. In this case, the worker's effort level conditional on accepting employment is $\tilde{e}$ (regardless of whether $w > \overline{w}(p_1)$ or $w \leq \overline{w}(p_1)$). Therefore, the firm's maximization problem is

$$\begin{aligned} \max_w \pi_1 + E\pi_2 &= \max_w \frac{3}{2}p_1 e(w,p_1) - 2w + \frac{1}{2}c(e(w,p_1)) \\ &= \max\Big\{0, \max_w \frac{3}{2}p_1\tilde{e}(w,p_1) - 2w + \frac{1}{2}c(\tilde{e}(w,p_1))\Big\}, \end{aligned}$$

where the 0 is the firm's outside-option payoff. The first-order condition gives

$$\left(\frac{3}{2}p_1 + \frac{1}{2}c'(\tilde{e}(w,p_1))\right)\frac{\partial \tilde{e}(w,p_1)}{\partial w} = 2.$$

Moreover, differentiating the equation defining the worker's solution $\tilde{e}(w,p_1)$, $\tilde{\pi}(w,\tilde{e};p_1) = \tilde{u}(w,\tilde{e})$, with respect to w gives

$$c'(\tilde{e}(w,p_1))\frac{\partial \tilde{e}(w,p_1)}{\partial w} + p_1\frac{\partial \tilde{e}(w,p_1)}{\partial w} = 2.$$

These two equations taken together imply $c'(\tilde{e}(w,p_1)) = p_1$, and so $\tilde{e}(w,p_1) = e_{\text{eff}}(p_1)$. The optimizing wage equates the surplus payoffs at the efficient level of effort and is given by equation (4).

Case 2. $\mu_A > \frac{1-2\sigma}{2}$. In this case, the solution to the worker's problem depends on whether $w > \overline{w}(p_1)$ or $w \leq \overline{w}(p_1)$. Hence the firm's maximization problem is now

$$\max\left\{0, \underbrace{\max_{w \leq \overline{w}(p_1)} p_1\tilde{e}(w,p_1) - w}_{\text{(A)}}, \underbrace{\max_{w > \overline{w}(p_1)} p_1 e_{\text{I}} - w}_{\text{(B)}}\right\}.$$

The candidate interior solution to sub-problem (A), w_{A}, is identical to the previous case with the solution being given by $w_{\text{A}} = \frac{c(e_{\text{eff}}(p_1)) + p_1 e_{\text{eff}}(p_1) - \pi_0 + u_0}{2}$.

As in the proof of Proposition 1, the maximand in (B) is strictly decreasing in w, but a candidate solution is that the worker's employment constraint binds with equality. However, even if there exists a solution to (B), $w_{\text{B}} > \overline{w}(p_1)$, the solution to (A) dominates it since, recalling that $\tilde{e}(\overline{w}(p_1),p_1) = e_{\text{I}}$,

$$p_1\tilde{e}(w_{\text{A}},p_1) - w_{\text{A}} \geq p_1\tilde{e}(\overline{w}(p_1),p_1) - \overline{w}(p_1) > p_1 e_{\text{I}} - w_{\text{B}}.$$

We conclude that the unique candidate solution is the solution to (A). Note that the solution to (A) is interior if and only if $w_{\text{A}} \leq \overline{w}(p_1)$, or equivalently, $e_{\text{I}} \geq \tilde{e}(w,p_1) = e_{\text{eff}}(p_1)$. Given equation (5), this is equivalent to $\mu_A \leq \frac{1-\sigma}{2}$, which is guaranteed by Assumption A′. Hence $(w_1^*, e_1^*) = (w_{\text{A}}, e_{\text{eff}}(p_1))$.

Finally, we check the firm's and the worker's employment constraints. The firm makes the wage offer whenever $\pi_1 + E\pi_2 \geq 0$. Note that

$$\begin{aligned}
\pi_1 + E\pi_2 &= \frac{3}{2}p_1e_1 - 2w_1 + \frac{1}{2}c(e_1) + \frac{1}{2}EM(p_2) \\
&= \frac{3}{2}p_1e_1 - \Big(c(e_1) + p_1e_1 - \pi_0 + u_0\Big) + \frac{1}{2}c(e_1) + \frac{1}{2}EM(p_2) \\
&= \frac{p_1e_1 - c(e_1) + EM(p_2)}{2} + \pi_0 - u_0 \\
&= \frac{M(p_1) + EM(p_2)}{2} + \pi_0 - u_0.
\end{aligned}$$

Therefore, the firm makes the wage offer if and only if

$$u_0 - \pi_0 \leq \frac{M(p_1) + EM(p_2)}{2}.$$

The worker accepts the job offer whenever $U_1^E + EU_2^E \geq 0$. Note that

$$\begin{aligned} U_1^E &= w_1 - c(e_1) - (1-\sigma)u_0 = \frac{p_1 e_1 - c(e_1) - \pi_0 + u_0}{2} - (1-\sigma)u_0 \\ &= \frac{M(p_1) - \pi_0 + (2\sigma - 1)u_0}{2} \end{aligned}$$

and

$$\begin{aligned} EU_2^E &= \frac{EM(p_2) - \pi_1 + (2\sigma-1)u_1}{2} = \frac{EM(p_2) - (p_1 e_1 - w_1) + (2\sigma-1)(w_1 - c(e_1))}{2} \\ &= \frac{2EM(p_2) - (p_1 e_1 - c(e_1) + \pi_0 - u_0) + (2\sigma-1)(p_1 e_1 - c(e_1) - \pi_0 + u_0)}{4} \\ &= \frac{EM(p_2) - (1-\sigma)M(p_1) - \sigma\pi_0 + \sigma u_0}{2}. \end{aligned}$$

Summing, we have

$$U_1^E + EU_2^E = \frac{\sigma M(p_1) + EM(p_2) - (1+\sigma)\pi_0 + (3\sigma - 1)u_0}{2}.$$

Hence the worker accepts the offer if and only if

$$(1+\sigma)\pi_0 - (3\sigma - 1)u_0 \leq \sigma M(p_1) + EM(p_2).$$

Both of these conditions are guaranteed by Assumption B$'$.

□

Proposition 4. *At the equilibrium described in Proposition 3: for $t = 1, 2$,*

1. *w_t^* and e_t^* are both increasing in p_t.*
2. *e_t^* does not depend on u_0 nor π_0.*
3. *w_t^* is increasing in u_0 and decreasing in π_0.*

Proof: The result has already been proven for $t = 2$ in Proposition 2. Since w_1^* is given by equation (4), it is increasing in p_1 and u_0, and decreasing in π_0. On the other hand, $e_1^* = e_{\text{eff}}(p_1)$, and so it is independent of both u_0 and π_0, and is increasing in p_1.

□

Lemma 2. *Under Assumption A, if $\lambda > 1$, then for any $p > 1$, there exists a $\overline{w}(p)$ such that for $w_0 < \overline{w}(p)$:*

1. *Effort responds more to wage cuts than to wage increases:* $\frac{\lim_{w\uparrow w_0}\frac{\partial e(w,p)}{\partial w}}{\lim_{w\downarrow w_0}\frac{\partial e(w,p)}{\partial w}} > 1$.

2. *Effort is more responsive to wage changes when effort is below the reference level of effort than when effort is below it:* $\frac{\lim_{e\uparrow e_0}\frac{\partial e(w,p)}{\partial w}}{\lim_{e\downarrow e_0}\frac{\partial e(w,p)}{\partial w}} > 1$.

If $w_0 \geq \overline{w}(p)$*, then* $e(w,p)$ *is constant in* w*.*

Proof: By an analogous argument as in the proof of Lemma 1, on $w < \overline{w}(p)$, $e(w,p) \equiv \tilde{e}(w,p)$, where $\tilde{e}(w,p)$ is defined by $\tilde{u}(w,\tilde{e}) = \tilde{\pi}(w,\tilde{e})$. Differentiating with respect to w:

$$\frac{\partial e}{\partial w} = \begin{cases} \frac{2}{p+c'(e)} & w \geq w_0, e \leq e_0, \\ \frac{2}{p+\lambda c'(e)} & w \geq w_0, e > e_0, \\ \frac{1+\lambda}{p+c'(e)} & w < w_0, e \leq e_0, \\ \frac{1+\lambda}{p+\lambda c'(e)} & w < w_0, e > e_0. \end{cases}$$

Since $\lambda > 1$ and $c' > 0$, the results of the lemma follow immediately. For $w_0 \geq \overline{w}(p)$, a similar argument as in the proof of Lemma 1 implies that $e(w,p)$ is constant in w.

□

Proposition 5. *Under Assumption A, if the firm hires the worker in equilibrium, then the equilibrium* (w^*, e^*) *is unique almost surely. Moreover:*

1. *If* $p \in \left(\underline{p}_{\text{w-rigid}}, \overline{p}_{\text{w-rigid}}\right)$*, then* $w^* = w_0$*,* $e^* > e_0$*, and* e^* *is strictly decreasing in* p*.*

2. *If* $p \in \left(\underline{p}_{\text{e-rigid}}, \overline{p}_{\text{e-rigid}}\right)$*, then* $e^* = e_0$*,* $w^* > w_0$*, and* w^* *is strictly increasing in* p*.*

3. *If* p *is outside the above ranges, then* w^* *and* e^* *are both strictly increasing in* p*.*

Proof: Consider the steady-state case: $c'(e_0) = p_0$. If $p = p_0$, then clearly $w^* = w_0$ and $e^* = e_0$.

For $p < p_0$, the firm has two options: $w < w_0$ or $w \geq w_0$. Consider first $w \geq w_0$. Since the worker chooses effort to equalize the surpluses, it must be that $e > e_0$ since $p < p_0$ and $w \geq w_0$. To see this, recall that the equalizing level of effort is given by

$$w - \theta c(e) - w_0 + \theta c(e_0) = pe - w - p_0 e_0 + w_0$$

for $\theta \in \{1, \lambda\}$, depending on whether or not $e > e_0$. Rearranging and using the fact that $w \geq w_0$ and $p_0 > p$,

$$\theta(c(e) - c(e_0)) = 2(w - w_0) + p_0 e_0 - pe > p_0(e_0 - e).$$

Since $c' > 0$, this can only hold for $e > e_0$. Now Lemma 2 implies that $\frac{\partial e}{\partial w} = \frac{2}{p+\lambda c'(e)}$. Note that the profit $\pi(w, e(w)) = pe - w$ is in fact decreasing in w since

$$\frac{\partial \pi}{\partial w} = p\frac{\partial e}{\partial w} - 1 = \frac{p - \lambda c'(e)}{p + \lambda c'(e)} < 0.$$

The final inequality follows from the fact that $c'(e) > c'(e_0) = p_0 > p$ and $\lambda \geq 1$. Thus the firm's solution on $w \geq w_0$ is $w^* = w_0$, and e^* is defined by

$$\lambda c(e^*) + pe^* = p_0 e_0 + \lambda c(e_0). \tag{6}$$

Turning to $w < w_0$, we now have $e < e_0$ using a similar argument as before. By Lemma 2, $\frac{\partial e}{\partial w} = \frac{1+\lambda}{p+c'(e)}$. Therefore, $\frac{\partial \pi}{\partial w} = \frac{\lambda p - c'(e)}{p+c'(e)}$. The first-order condition implies $c'(e^*) = \lambda p$. However, we require $e < e_0$, or equivalently, $c'(e) < c'(e_0) = p_0$. Therefore, for $p \in \left(\frac{p_0}{\lambda}, p_0\right)$, the maximization problem has no solution. If $p \leq \frac{p_0}{\lambda}$, then the solution is given by $c'(e^*) = \lambda p$, and w^* is the wage such that the surpluses are equalized:

$$w^* = \frac{pe^* - p_0 e_0 + (1+\lambda)w_0 + c(e^*) - c(e_0)}{1+\lambda}.$$

Since $c'(e^*) = \lambda p$, we know that $e^* < e_0$, and thus $w^* = w_0 + \frac{pe^* - p_0 e_0 + c(e^*) - c(e_0)}{1+\lambda} < w_0$.

We conclude as follows: in the steady-state case with $p \in \left(\frac{p_0}{\lambda}, p_0\right)$, $w^* = w_0$ and $e^* > e_0$ is defined by equation (6). Analysis of equation (6) shows that e^* must be decreasing in p since the right-hand side is constant and the left-hand side is strictly increasing in e. For $p < \frac{p_0}{\lambda}$, $e^* < e_0$ and $w^* < w_0$ are strictly increasing in p. Denoting profit when the firm sets that wage by $\pi_{<w_0}(p)$ and profit when the firm sets wage w_0 by $\pi_{w_0}(p)$, note that because both functions are continuous in p, $\pi_{w_0}\left(\frac{p_0}{\lambda}\right) = \pi_{<w_0}\left(\frac{p_0}{\lambda}\right)$. The firm is therefore indifferent, and thus the equilibrium is non-unique, when the price realization is exactly $\frac{p_0}{\lambda}$.

For $p > p_0$, the firm again has the same two options: $w < w_0$ or $w \geq w_0$. Consider first $w < w_0$. As before, it must be the case that $e < e_0$. Therefore, $\frac{\partial e}{\partial w} = \frac{1+\lambda}{p+c'(e)}$. Hence, $\frac{\partial \pi}{\partial w} = \frac{p\lambda - c'(e)}{p+c'(e)} > 0$ since $e < e_0$ implies $c'(e) < c'(e_0) = p_0 < p$ and $\lambda \geq 1$. That is, profit is strictly increasing in wage and so there is no solution for $w < w_0$.

Now consider $w \geq w_0$. We analyze three subcases: (a) $w = w_0$, (b) $w > w_0$ and $e \leq e_0$, and (c) $w > w_0$ and $e > e_0$. For (a), since $p > p_0$, it must be that $e < e_0$ since the worker chooses effort to equalize the surplus payoffs. Thus for both (a) and (b), we have that $\frac{\partial \pi}{\partial w} = \frac{p - c'(e)}{p+c'(e)} > 0$ since $c'(e) < c'(e_0) = p_0 < p$. But this implies that $e > e_0$, a contradiction. Thus we know that if $w \geq w_0$, then we are in case (c).

For (c), $\frac{\partial \pi}{\partial w} = \frac{p-\lambda c'(e)}{p+\lambda c'(e)}$. The firm's first-order condition thus implies $c'(e) = \frac{p}{\lambda}$. Since we are on the domain $e > e_0$, or equivalently, $c'(e) > c'(e_0) = p_0$, for $p < \lambda p_0$, we have a corner solution, $e^* = e_0$. Since the wage is such that the surpluses are equalized, it must be increasing in p since profit is increasing in p, and effort and thus utility are otherwise constant. For $p \geq \lambda p_0$, the solution is given by $c'(e^*) = \frac{p}{\lambda}$ and the wage w^* that equalizes the surpluses:

$$w^* = \frac{pe^* - p_0 e_0 + 2w_0 + \lambda c(e^*) - \lambda c(e_0)}{2}.$$

Clearly e^* and w^* are both increasing in p since $c' > 0$ and $c'' > 0$.

Now consider the recent-increase case: $p_0 = \lambda c'(e_0)$. Define $p_{00} \equiv \frac{p_0}{\lambda}$. Then the previous analysis applies to $p_{00} = c'(e_0)$. That is, for $p \in \left(\frac{p_{00}}{\lambda}, p_{00}\right) = \left(\frac{p_0}{\lambda^2}, \frac{p_0}{\lambda}\right)$, $w^* = w_0$, and $e^* > e_0$ is decreasing in p. For $p \in (p_{00}, \lambda p_{00}) = \left(\frac{p_0}{\lambda}, p_0\right)$, $e^* = e_0$, and $w^* \geq w_0$ is increasing in p. For all other p, w^* and e^* are increasing in p.

Finally, the recent-decrease case: $\lambda p_0 = c'(e_0)$. Similarly to above, define $p_{00} \equiv \lambda p_0$, and apply the steady-state-case result to $p_{00} = c'(e_0)$. That is, for $p \in \left(\frac{p_{00}}{\lambda}, p_{00}\right) = (p_0, \lambda p_0)$, $w^* = w_0$, and $e^* > e_0$ is decreasing in p. For $p \in (p_{00}, \lambda p_{00}) = (\lambda p_0, \lambda^2 p_0)$, $e^* = e_0$, and $w^* \geq w_0$ is increasing in p. For all other p, w^* and e^* are increasing in p.

□

Proposition 6. *Under Assumptions A and B, if the worker is not loss averse ($\lambda = 1$), then for any $p > 1$, the equilibrium transaction is UPE and MPE.*

Proof: By Proposition 1, $e^* = \tilde{e}(w^*, e^*) = e_{\text{eff}}(p)$. Therefore, the equilibrium is MPE. Since we are in the case of no loss aversion,

$$U^E = \mu(w - c(e)) + (1 - \mu)(pe - w) + \sigma u_0,$$

where $\mu \in \{\mu_D, \mu_A\}$. Thus for each μ,

$$\frac{\partial U/\partial w}{\partial U/\partial e} = \frac{\mu - (1-\mu)}{-\mu c'(e) + (1-\mu)p}.$$

Since $c'(e^*) = c'(e_{\text{eff}}(p)) = p$,

$$\lim_{(w,e)\to(w^*,e^*)} \frac{\partial U/\partial w}{\partial U/\partial e} = \frac{2\mu - 1}{(1-2\mu)p} = \frac{-1}{p}.$$

Moreover, since $\pi = pe - w$,

$$\frac{\partial \pi/\partial w}{\partial \pi/\partial e} = \frac{-1}{p}.$$

Since $\frac{\partial U/\partial w}{\partial U/\partial e} = \frac{\partial \pi/\partial w}{\partial \pi/\partial e}$, (w^*, e^*) is UPE.

□

Proposition 7. *Under Assumption A, if the worker is loss averse ($\lambda > 1$) and the firm hires the worker in equilibrium, then the equilibrium transaction is UPE if and only if* $p \notin \left(\underline{p}_{\text{w-rigid}}, \overline{p}_{\text{w-rigid}}\right) \cup \left(\underline{p}_{\text{e-rigid}}, \overline{p}_{\text{e-rigid}}\right)$. *The equilibrium transaction is also MPE if and only if $p = \overline{p}_{\text{w-rigid}}$, or equivalently, if and only if the equilibrium transaction (w^*, e^*) satisfies $w^* = w_0$ and $e^* = e_0$.*

Proof: We first show that (w^*, e^*) is not UPE for $p \in (\underline{p}_{\text{w-rigid}}, \overline{p}_{\text{w-rigid}}) = \left(\frac{p_{00}}{\lambda}, p_{00}\right)$. Proposition 5 implies that $w^* = w_0$, $e^* > e_0$, and $c'(e^*) > c'(e_0) = p_{00} > p$. Let $\Delta > 0$ be small enough so that $e' = e^* - \Delta > e_0$ and define $(w', e') = (w^* - p\Delta, e^* - \Delta)$ so that

$$\pi(w', e') = pe' - w' = p(e^* - \Delta) - (w^* - p\Delta) = pe^* - w^* = \pi(w^*, e^*).$$

Since c is convex and $c'(e') > c'(e_0) > p$,

$$c(e^*) = c(e' + \Delta) \geq c(e') + c'(e')\Delta > c(e') + p\Delta.$$

Therefore, the worker's change in wage is $w' - w^* = -p\Delta$, and the change in the cost of effort is $c(e') - c(e^*) < -p\Delta$. It follows that $u(w', e') - u(w^*, e^*) > 0$. Since the firm's profit is unchanged, it must be that $U(w', e') > U(w^*, e^*)$. Thus (w^*, e^*) is not UPE.

We now show that (w^*, e^*) is not UPE for $p \in (\underline{p}_{\text{e-rigid}}, \overline{p}_{\text{e-rigid}}) = (p_{00}, \lambda p_{00})$. Proposition 5 implies that $e^* = e_0$ and $w^* > w_0$. Let $\Delta > 0$ be small enough so that $c'(e^* + \Delta) < p$ (this is possible since $c'(e^*) = p_{00} < p$ and c' is continuous), and define $(w', e') = (w^* + p\Delta, e^* + \Delta)$ so that $\pi(w', e') = \pi(w^*, e^*)$. Again, since c is convex,

$$c(e^*) = c(e' - \Delta) > c(e') - c'(e')\Delta > c(e') - p\Delta$$

since $c'(e') < p$. Thus, the worker's change in wage is $w' - w^* = p\Delta$, and his change in the cost of effort is $c(e') - c(e^*) < p\Delta$. We then have that $u(w', e') - u(w^*, e^*) > 0$. Since the firm's profit is unchanged, it must be that $U(w', e') > U(w^*, e^*)$. Thus (w^*, e^*) is not UPE.

There are three more cases to consider: (1) $p > \overline{p}_{\text{e-rigid}}$, (2) $p < \underline{p}_{\text{w-rigid}}$, and (3) $p = \overline{p}_{\text{w-rigid}}$. We analyze each in turn.

Case (1). From the proof of Proposition 5, we know that in this case, $w^* > w_0$, $e^* > e_0$, and $c'(e^*) = p/\lambda$. Therefore,

$$\frac{\partial U/\partial w}{\partial U/\partial e} = \frac{2\mu - 1}{-\mu\lambda c'(e^*) + p(1-\mu)} = \frac{2\mu - 1}{-\mu p + p(1-\mu)} = -\frac{1}{p}.$$

Since (as in the proof of Proposition 6) $\frac{\partial \pi / \partial w}{\partial \pi / \partial e} = -\frac{1}{p}$ as well, (w^*, e^*) is UPE.

Case (2). From the proof of Proposition 5, we know that in this case, $w^* < w_0$, $e^* < e_0$, and $c'(e^*) = \lambda p$. Therefore,

$$\frac{\partial U / \partial w}{\partial U / \partial e} = \frac{(1+\lambda)\mu - 1}{-\mu c'(e^*) + p(1-\mu)} = \frac{(1+\lambda)\mu - 1}{-\mu\lambda p + p(1-\mu)} = -\frac{1}{p}.$$

Thus, as before, (w^*, e^*) is UPE.

Case (3). From the proof of Proposition 5, we know that in this case, $w^* = w_0$, $e^* = e_0$, and $c'(e_0) = p$. For contradiction, suppose there exists (w', e') that utility-Pareto dominates (w^*, e^*). Define Δ_e and Δ_w by $e' = e_0 + \Delta_e$ and $w' = w_0 + \Delta_w$. Since $\pi(w', e') \geq \pi(w_0, e_0)$, we know that $\Delta_w \leq p\Delta_e$, and since (w', e') is a utility-Pareto improvement, we know that $\text{sgn}(\Delta_w) = \text{sgn}(\Delta_e)$.

Below, we will use the following result: Since c is convex,

$$c(e') = c(e_0 + \Delta_e) \geq c(e_0) + c'(e_0)\Delta_e = c(e_0) + p\Delta_e.$$

There are three sub-cases to consider: (a) $\Delta_e > 0$, (b) $\Delta_e < 0$, and (c) $\Delta_e = 0$. Starting with sub-case (a), the worker is experiencing a loss in the effort domain since $\Delta_e > 0$ and a gain in the wage domain since $\Delta_w > 0$. Thus,

$$\begin{aligned}
\tilde{u}(w', e') - \tilde{\pi}(w', e') &= 2w' - 2w_0 - pe' + pe_0 - \lambda c(e') + \lambda c(e_0) \\
&\leq 2(w_0 + \Delta_w) - 2w_0 - p(e_0 + \Delta_e) + pe_0 - \lambda(c(e_0) + p\Delta_e) + \lambda c(e_0) \\
&= 2\Delta_w - p\Delta_e - \lambda p\Delta_e \leq (2 - (1+\lambda))p\Delta_e < 0,
\end{aligned}$$

and therefore the worker is in the region of disadvantageous unfairness. Dropping the u_0 and π_0 terms in the worker's utility,

$$\begin{aligned}
U(w', e') &= \mu_D \left(w' - \lambda c(e')\right) + (1 - \mu_D)\left(pe' - w'\right) \\
&= (2\mu_D - 1)w' - \lambda\mu_D c(e') + (1 - \mu_D)pe' \\
&\leq (2\mu_D - 1)(w_0 + \Delta_w) - \lambda\mu_D(c(e_0) + p\Delta_e) + (1 - \mu_D)p(e_0 + \Delta_e) \\
&< (2\mu_D - 1)w_0 - \mu_D c(e_0) + (1 - \mu_D)pe_0 + (2\mu_D - 1)\Delta_w + (1 - \mu_D - \lambda\mu_D)p\Delta_e \\
&< (2\mu_D - 1)w_0 - \mu_D c(e_0) + (1 - \mu_D)pe_0 \\
&= U(w_0, e_0).
\end{aligned}$$

But this contradicts the hypothesis that (w', e') utility-Pareto dominates (w_0, e_0).

Now consider sub-case (b), $\Delta_e < 0$. Since $\Delta_w < 0$, the worker is experiencing a loss in the wage domain. Since $c(e') - c(e_0) \geq p\Delta_e \geq \Delta_w$, we know that $u(w', e') - u(w_0, e_0) = \Delta_w - (c(e') - c(e_0)) \leq$

0, with strict inequality if $\Delta_w < p\Delta_e$. Suppose first that $\Delta_w = p\Delta_e$. Then $\pi(w', e') = \pi(w_0, e_0)$, and so $\tilde{\pi}(w', e') = \tilde{\pi}(w_0, e_0)$. Thus, for $\mu \in \{\mu_A, \mu_D\}$,

$$\begin{aligned} U(w', e') &= \mu\tilde{u}(w', e') + (1-\mu)\tilde{\pi}(w', e') + \sigma u_0 \\ &= \mu(u(w', e') - u_0) + (1-\mu)\tilde{\pi}(w_0, e_0) + \sigma u_0 \\ &\leq \mu(u(w_0, e_0) - u_0) + (1-\mu)\tilde{\pi}(w_0, e_0) + \sigma u_0 \\ &= \mu\tilde{u}(w_0, e_0) + (1-\mu)\tilde{\pi}(w_0, e_0) + \sigma u_0 = U(w_0, e_0). \end{aligned}$$

Therefore (w', e') cannot utility-Pareto dominate (w_0, e_0), a contradiction. Now suppose $\Delta_w < p\Delta_e$. Then since $\pi(w', e') > \pi(w_0, e_0)$, $u(w', e') < u(w_0, e_0)$, and (w_0, e_0) is a surplus-equalizing transaction, it must be that the worker is in the disadvantageously-unfair region. Furthermore, $\pi(w', e') > \pi(w_0, e_0)$ implies that $\tilde{\pi}(w', e') > \tilde{\pi}(w_0, e_0)$. Since $\mu_D \geq 1$ (from A(i)),

$$\begin{aligned} U(w', e') &= \mu_D\tilde{u}(w', e') + (1-\mu_D)\tilde{\pi}(w', e') + \sigma u_0 \\ &\leq \mu_D(u(w', e') - u_0) + (1-\mu_D)\tilde{\pi}(w_0, e_0) + \sigma u_0 \\ &< \mu_D(u(w_0, e_0) - u_0) + (1-\mu_D)\tilde{\pi}(w_0, e_0) + \sigma u_0 \\ &= \mu_D\tilde{u}(w_0, e_0) + (1-\mu_D)\tilde{\pi}(w_0, e_0) + \sigma u_0 = U(w_0, e_0). \end{aligned}$$

Again, this contradicts the hypothesis that (w', e') utility-Pareto dominates (w_0, e_0).

Last, we turn to sub-case (c), $\Delta_e = 0$. Since $\Delta_w \leq p\Delta_e$, we know that either $\Delta_w < 0$ or $\Delta_w = 0$. But the latter is impossible because then profit and utility would be unchanged (contradicting UPE), so $\Delta_w < 0$. Since $\Delta_e = 0$ and $\Delta_w < 0$, the worker is in the region of disadvantageous unfairness. But then a decrease in the wage reduces utility, so $U(w', e') < U(w_0, e_0)$, contradicting the hypothesis that (w', e') utility-Pareto dominates (w_0, e_0). We conclude that in case (3), $(w^*, e^*) = (w_0, e_0)$ is UPE.

Finally, we consider when the equilibrium is MPE, i.e., when $c'(e) = p$. According to Proposition 5, this only occurs when $p = p_{00}$ and $(w^*, e^*) = (w_0, e_0)$.

□