

Kfir Eliaz and Ran Spiegler

Managing intrinsic motivation in a long-run relationship

Discussion paper

Original citation:

Eliaz, Kfir and Spiegler, Ran (2014) *Managing intrinsic motivation in a long-run relationship*. CFM discussion paper series, CFM-DP2014-14. Centre For Macroeconomics, London, UK.

Originally available from the [Centre For Macroeconomics](#)

Financial support from the Sapir Centre and ERC grant no. 230251 is gratefully acknowledged.

This version available at: <http://eprints.lse.ac.uk/58020/>

Available in LSE Research Online: July 2014

© 2014 The Authors

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

Managing Intrinsic Motivation in a Long-Run Relationship*

Kfir Eliaz[†] and Ran Spiegler[‡]

June 23, 2014

Abstract

We study a repeated principal-agent interaction, in which the principal offers a "spot" wage contract at every period, and the agent's outside option follows a Markov process with *i.i.d* shocks. If the agent rejects an offer, the two parties are permanently separated. At any period during the relationship, the agent is productive if and only if his wage does not fall below a "reference point" (by more than an infinitesimal amount), which is defined as his lagged-expected wage in that period. We characterize the game's unique subgame perfect equilibrium. The equilibrium path exhibits an aspect of wage rigidity. The agent's total discounted rent is equal to the maximal shock value.

1 Introduction

The standard principal-agent model is built on the premise that the agent needs to be incentivized in order to exert effort on a task. This requires the principal to condition the agent's wage on a verifiable signal of his effort. However, in many environments such information is either unavailable or very imprecise, which forces the principal to rely on the agent's "intrinsic motivation". For instance, when a parent hires a nanny, many of the effects of good care are unobservable. Similarly, if one wishes to implement effective care by a surgeon, forcing her to stick to some protocol will often miss the target, and the effects of a good surgery are confounded with other factors and therefore hard to contract on. Even when it is relatively easy to check whether the task itself was completed successfully - say, flying a plane from point A to point B on schedule - it may be difficult to discern whether disruptions in performance could have been

*We thank Yair Antler, Alex Frug and Eeva Mauring for helpful comments. Financial support from the Sapir Centre and ERC grant no. 230251 is gratefully acknowledged.

[†]Tel Aviv University and University of Michigan. E-mail: kfire@post.tau.ac.il.

[‡]Tel Aviv University, University College London and CFM. E-mail: rani@post.tau.ac.il.

avoided (e.g., unnecessarily delaying a flight due to a malfunction that the pilot could have fixed himself).¹

Intrinsic motivation is inherently a *dynamic* property - an agent who is initially motivated may temporarily *lose* his motivation in the course of his relationship with the principal. In particular, numerous studies in the literature - notably, Akerlof (1982), Akerlof and Yellen (1990), Bewley (1999), Fehr et al. (2009) - have argued that intrinsic motivation is *reference-dependent*. An agent may become demotivated when his compensation falls below his expectations. This means that temporal variations in the agent's compensation, which may reflect variations in his outside option, can adversely affect the agent's motivation. Hence, in situations with limited contractual instruments, the principal is faced with the problem of optimally managing the agent's motivation: trading-off the cost and benefit of keeping the agent motivated.

To study this dynamic principal-agent problem, we propose a simple model in which the agent is motivated whenever his current wage does not fall significantly below some reference point. The agent enters every period in the principal-agent relationship with some "reference wage" e_t , and the principal makes a "spot" wage offer w_t . That is, the principal is unable to commit to a wage strategy ex-ante, and he is unable to condition the wage on any verifiable signal. If the agent declines, the two parties are permanently separated, and the agent receives at every period $s \geq t$ a publicly observed outside payment θ_t , which evolves according to some Markov process. Both parties maximize their discounted expected payoffs. If the agent accepts the offer, his output is reference-dependent. His output is 1 if $w_t \geq e_t - \lambda$ (where $\lambda > 0$) and 0 otherwise. Our entire analysis focuses on the $\lambda \rightarrow 0$ limit.

To complete the model, we need to specify the rule that governs the evolution of the reference wage e_t . Inspired by Kőszegi and Rabin (2006), we assume that at every t , e_t is equal to the agent's *lagged-expected wage* - that is, the "rational" expectation of w_t calculated at the end of period $t - 1$ according to the parties' continuation strategies, and conditional on the event that the relationship is not severed at t . The expectation-based component of the reference captures the idea that a wage is treated as a disappointment or as a pleasant surprise, depending on how it compares with the agent's former expectations: even when the wage rises relative to the previous period, the agent may still be disappointed if he expected a big raise. The lagged-expectations aspect of the reference point captures the idea that reference points are sluggish in adapting to new information. In this respect, reference points are like habits, which

¹See the post "Incomplete Contracts and Labor Disputes" at <http://cheaptalk.org/2012/10/04/incomplete-contracts-and-labor-disputes/>.

often take time to change in the face of new circumstances.

Our task is to characterize subgame perfect equilibria in this game. In particular, we want to understand the implications of the agent's reference-dependent intrinsic motivation on the structure of the equilibrium wage and the agent's rent. To illustrate the possible effects of reference dependence, consider perfectly myopic parties. This means that the agent's participation wage at period t is θ_t . Assume that θ_t can take two values, $\underline{\theta}$ and $\bar{\theta}$, with equal probability (independently of the history), where $\underline{\theta} < \bar{\theta} < 1$. Suppose that in equilibrium the parties' relationship is not severed at t for any realization of θ_t . Let $w(\theta)$ denote the equilibrium wage when $\theta_t = \theta$. Then, $e_t = \frac{1}{2}w(\underline{\theta}) + \frac{1}{2}w(\bar{\theta})$. If the principal paid the agent his reference wage in equilibrium, we would have $e = \frac{1}{2}\underline{\theta} + \frac{1}{2}\bar{\theta} > \underline{\theta}$. If λ is small, the agent will produce zero output when $\theta_t = \underline{\theta}$. Therefore, it would be profitable for the principal to deviate to $w_t = e$ in the state $\underline{\theta}$. In fact, the only wage strategy that is consistent with equilibrium in the $\lambda \rightarrow 0$ limit is $w(\underline{\theta}) = w(\bar{\theta}) = \bar{\theta}$.

The equilibrium strategy in this example has two noteworthy features: (i) *wage rigidity* - the wage is invariant to the fluctuations in the agent's outside option; (ii) *efficiency wages* - the principal pays the agent a wage above the reservation level in order to ensure high output. The example thus naturally links the two phenomena together. Note that the efficiency wage effect does not arise from a conventional incentive constraint, but from the desire not to disappoint the agent and rob him of his intrinsic motivation.

When parties are not myopic, a new consideration arises: the efficiency-wage effect means that the agent expects to earn rents in the future, and this lowers his current reservation point. Since this wage in turn determines the equilibrium reference wage, finding the equilibrium wage strategy requires us to find a fixed point of a coupled pair of functional equations: the dynamic reservation-wage equation after every history, and the equation that defines the reference wage after every history. From a technical point of view, this novel fixed-point constitutes the paper's core.

This note follows up Eliaz and Spiegler (2013), which essentially embedded a more elaborate version of the above myopic case in a search-matching model of the labor market (effective myopia arose from a short horizon of the employment relation, rather than from a zero discount factor), and derived implications for patterns of wage rigidity, endogenous job destruction and enhanced volatility of labor-market tightness. The technical challenge in Eliaz and Spiegler (2013) arose from the assumption that when workers contemplate declining an offer, they anticipate the possibility of being re-matched with a new employer in the future. The present note abstracts from re-

matching and focuses instead on the pure principal-agent problem, and on the new considerations that arise from the infinite horizon of the principal-agent relationship. Re-incorporating a model along these lines in a larger search-matching model of the labor market is an important challenge for future research.

The idea that an agent's productivity in a task depends directly on his beliefs has precedents in the theoretical literature. Compte and Postlewaite (2004) analyze a model in which the agent's ability to complete a task depends on his subjective belief regarding this ability (namely, his "confidence"). They show that in such a model, biased beliefs can enhance welfare. Similarly, Fang and Moscarini (2005) show that when an agent's performance depends on his confidence, an informed principal may prefer to design a wage scheme that does not differentiate between abilities.

2 A Model

Two players, referred to as a principal and an agent, play a discrete time, infinite-horizon game with perfect information. At the beginning of every period $t = 1, 2, \dots$, the principal makes a wage offer $w_t \in \mathbb{R}$. If the agent rejects the offer, the relationship is terminated, and the agent (principal) collects a payoff of θ_s (0) at every period $s \geq t$. We assume that $\theta_t = \Psi(\theta_{t-1}) + \varepsilon_t$, where Ψ is a deterministic function and ε_t is *i.i.d* according to a *cdf* F with mean zero. Let $\bar{\varepsilon}$ denote the highest value that ε_t can take. We assume that Ψ and F are such that θ_t always takes values in $(0, 1)$. If the agent accepts the offer at period t , he collects a payoff w_t , and the principal's payoff is $y_t = \mathbf{1}(w_t \geq e_t - \lambda) - w_t$, where $\lambda > 0$ and e_t is the agent's *reference point* at period t . We refer to y_t as the agent's output in period t . Both parties maximize discounted expected payoffs, with a discount factor $\delta \in [0, 1)$.

For every period t in which the agent is employed, let h_t denote the history of realized wages, output and the outside option up to and including period t , i.e. $h_t = (w_s, y_s, \theta_s)_{s=1}^t$. The history is commonly observed by both players. However, the agent's output is unverifiable, which is why the principal cannot condition the agent's wage on his output. A strategy for the principal is a function w that specifies a wage offer for every history h_{t-1} and realized outside option θ_t . A strategy for an agent is a function a that specifies for every (h_{t-1}, θ_t) and wage offer w_t a binary decision: "accept" ($a = 1$) or "reject" ($a = 0$).

To complete the description of the game, we need to specify how e_t is formed. Inspired by Kőszegi and Rabin (2006), we assume that it is the agent's lagged-expected wage at period t . More precisely, consider a history at the end of period $t - 1$ (i.e.,

before θ_t is realized), and fix the parties' continuation strategies from period t onwards. Then, e_t is the expectation of w_t , calculated according to these continuation strategies at the end of the period- $(t-1)$ history, conditional on the event that the agent accepts the principal's offer at period t . (If this is a null event, we set $e_t = 0$.) Thus, e_t - and therefore the principal's payoff at period t - is a function of the expectations that players hold at the end of period $t-1$. In equilibrium, these expectations will be correct. Given a strategy pair (w, a) , we let e denote the function that assigns for every history h_{t-1} a reference wage for period t .

Since the principal's payoff is defined in terms of the players' beliefs, this is not strictly speaking an extensive game, but an extensive *psychological game*, in the sense of Geanakoplos, Pearce and Stachetti (1989). However, since the belief-dependence is straightforward, we will work with the usual and familiar subgame perfect equilibrium (SPE) concept, which can be defined in terms of the usual single-deviation property: in equilibrium, each player's action at every history maximizes his discounted expected payoffs, given the continuation strategies of both players. More formally, an SPE in our game is a triple (w, a, e) that satisfies the following properties after every history (h_{t-1}, θ_t) . First, given (w, a, e) , the wage $w(h_{t-1}, \theta_t)$ maximizes the principal's discounted sum of expected payoffs. Second, for every wage offer w_t , the decision $a(h_{t-1}, \theta_t, w_t)$ maximizes the agent's discounted sum of expected payoffs. Third, given the principal's strategy w and the agent's strategy a , the reference function e satisfies

$$e(h_{t-1}) = \mathbb{E}[w(h_{t-1}, \theta_t) | a(h_{t-1}, \theta_t, w(h_{t-1}, \theta_t)) = 1]$$

and $e(h_{t-1}) = 0$ if the event $\{\theta_t | a(h_{t-1}, \theta_t, w(h_{t-1}, \theta_t)) = 1\}$ is null.

3 Analysis

Let us first consider a reference-independent benchmark model, in which the agent's output is always 1, independently of the history. (In other words, set $\lambda = \infty$.)

Claim 1 *Let $\lambda = \infty$. Then, there is a unique subgame perfect equilibrium: the agent's accepts any $w_t \geq \theta_t$ at every period t , and the principal offers $w_t = \theta_t$ at every t , independently of the history.*

This is a standard result due to the principal having all the bargaining power. Therefore, the proof is omitted. The equilibrium wage is entirely flexible and the agent earns no rent in equilibrium.

We now provide a characterization of subgame perfect equilibrium in the $\lambda \rightarrow 0$ limit, where the agent becomes unproductive whenever the actual wage falls below his reference point, however slightly.

Theorem 1 *There exists a unique SPE in the $\lambda \rightarrow 0$ limit. At every period t , following any history:*

- (i) *The principal offers $w_t = \Psi(\theta_{t-1}) + (1 - \delta)\bar{\varepsilon}$.*
- (ii) *The agent accepts any $w_t \geq \theta_t - \delta\bar{\varepsilon}$.*

Proof. Throughout the proof, we use h_{t-1} to denote a history $(\theta_s, w_s)_{s=1, \dots, t-1}$, where θ_s is the realized outside option in period s and w_s is the wage offer that the principal made in period s , such that the agent accepted all wage offers up to period $t - 1$. We denote by (h_{t-1}, θ_t) the immediate concatenation of h_{t-1} , right after θ_t is realized. With slight abuse of notation, we use $F(\theta_{t+1} \mid \theta_t)$ to denote the *cdf* over θ_{t+1} conditional on θ_t . Finally, we denote the agent's reference point at period t following the history h_{t-1} by $e(h_{t-1})$. We will say that the equilibrium is Markovian if players' strategies at (h_{t-1}, θ_t) are purely a function of (θ_{t-1}, θ_t) . We will say that the principal has a unique equilibrium payoff if there is a function $V(\theta)$ such that in any subgame perfect equilibrium, the principal's payoff at the beginning of any period t , just before the realization of θ_t , is $V(\theta_{t-1})$.

Step 1: In subgame perfect equilibrium, for every (h_{t-1}, θ_t) there is a reservation wage $\bar{w}(h_{t-1}, \theta_t) < 1$, such that the agent accepts every $w_t \geq \bar{w}(h_{t-1}, \theta_t)$.

Proof: If the agent rejects an offer at t , his continuation payoff is $B(\theta_t) = \mathbb{E} \left[\sum_{s \geq t} \delta^{s-t} \theta_s \mid \theta_t \right]$. Recall that by assumption, $\theta_t < 1$. Therefore, the agent will strictly prefer to accept every $w_t \in (\theta_t, 1)$. Suppose that the agent accepts some $w_t \leq \theta_t$ after (h_{t-1}, θ_t) , but rejects some other $w'_t > w_t$. Given that the agent accepts w_t , his payoff must be weakly greater than $\max\{B(\theta_t), w_t + \delta \mathbb{E}(B(\theta_{t+1}) \mid \theta_t)\}$. If he rejects w'_t , his payoff is by definition $B(\theta_t)$. Therefore, deviating to accepting w'_t would necessarily generate a payoff of at least $w'_t + \delta \mathbb{E}(B(\theta_{t+1}) \mid \theta_t)$, which is strictly above $B(\theta_t)$, a contradiction. It follows that for every (h_{t-1}, θ_t) , there is a threshold $\bar{w}(h_{t-1}, \theta_t)$ such that the agent accepts every $w_t \geq \bar{w}(h_{t-1}, \theta_t)$.

Step 2: Suppose that the principal has a unique equilibrium payoff. Then,

$$w_t(h_{t-1}, \theta_t) = \max\{\bar{w}(h_{t-1}, \theta_t), e(h_{t-1}) - \lambda\}$$

following every history (h_{t-1}, θ_t) .

Proof: By the definition of unique equilibrium payoff, in any subgame perfect equilibrium, the principal's equilibrium payoff in any subgame following some history h_t (just before the realization of θ_t) is purely a function of θ_{t-1} . Suppose that $w_t(h_{t-1}, \theta_t) > \bar{w}(h_{t-1}, \theta_t)$ and $w_t(h_{t-1}, \theta_t) \neq e(h_{t-1}) - \lambda$ after some history (h_{t-1}, θ_t) . If the principal deviates to $w_t(h_{t-1}, \theta_t) - \eta$, where $\eta > 0$ is arbitrarily small, he will not change the agent's output at t , and since the principal's payoff in the continuation game will not change, the deviation has no implications for the principal's continuation payoff. Therefore, the deviation is profitable. It follows that if $w_t(h_{t-1}, \theta_t) \geq \bar{w}(h_{t-1}, \theta_t)$, then $w_t(h_{t-1}, \theta_t) \in \{\bar{w}(h_{t-1}, \theta_t), e(h_{t-1}) - \lambda\}$. By the definition of $e(h_{t-1})$ and the result that $\bar{w}(h_{t-1}, \theta_t) < 1$, it follows that $e(h_{t-1}) < 1$. Therefore, the principal will always choose $w_t(h_{t-1}, \theta_t) = \max\{\bar{w}(h_{t-1}, \theta_t), e(h_{t-1}) - \lambda\}$ after every (h_{t-1}, θ_t) , because this maximizes his period t payoff, without affecting his continuation payoff.

Step 3: Suppose that the principal has a unique equilibrium payoff. Then, in the $\lambda \rightarrow 0$ limit, the equilibrium is uniquely defined as in the statement of the theorem.

Proof: Assume the premise holds. Then, by Step 2 and the definition of the reference wage:

$$e(h_{t-1}) = \int_{\theta_t} \max\{\bar{w}(h_{t-1}, \theta_t), e(h_{t-1}) - \lambda\} dF(\theta_t | \theta_{t-1})$$

In the $\lambda \rightarrow 0$ limit, the solution to this equation is

$$e(h_{t-1}) = \max_{\theta_t | \theta_{t-1}} \bar{w}(h_{t-1}, \theta_t) \tag{1}$$

By Step 2, the principal pays $w_t = e(h_{t-1})$ after every history (h_{t-1}, θ_t) , and by the definition of \bar{w} the agent always accepts this offer. The agent's participation wage $\bar{w}(h_{t-1}, \theta_t)$ is the wage that makes him indifferent between accepting and rejecting an offer following (h_{t-1}, θ_t) :

$$\bar{w}(h_{t-1}, \theta_t) + \mathbb{E} \left[\left(\sum_{s>t} \delta^{s-t} \max_{\theta_s | \theta_{s-1}} \bar{w}(h_{s-1}, \theta_s) \right) \mid \theta_t \right] = B(\theta_t)$$

This can be rewritten as

$$\begin{aligned}\bar{w}(h_{t-1}, \theta_t) &= B(\theta_t) - \delta \max_{\theta_{t+1}|\theta_t} \bar{w}(h_t, \theta_{t+1}) - \delta \mathbb{E} \left[\left(\sum_{s>t+1} \delta^{s-t} \max_{\theta_s|\theta_{s-1}} \bar{w}(h_{s-1}, \theta_s) \right) \mid \theta_t \right] \\ &= B(\theta_t) - \delta \max_{\theta_{t+1}|\theta_t} \bar{w}(h_t, \theta_{t+1}) - \delta \left[\mathbb{E}B(\theta_{t+1}) - \int_{\theta_{t+1}} \bar{w}(h_t, \theta_{t+1}) dF(\theta_{t+1} \mid \theta_t) \right]\end{aligned}$$

which is simplified into the recursive functional equation

$$\bar{w}(h_{t-1}, \theta_t) = \theta_t - \delta \left[\max_{\theta_{t+1}|\theta_t} \bar{w}(h_t, \theta_{t+1}) - \int_{\theta_{t+1}} \bar{w}(h_t, \theta_{t+1}) dF(\theta_{t+1} \mid \theta_t) \right]$$

or

$$\bar{w}(h_{t-1}, \theta_t) = \theta_t + \delta \mathbb{E} [\bar{w}(h_t, \theta_{t+1}) \mid \theta_t] - \delta \max_{\theta_{t+1}|\theta_t} \bar{w}(h_t, \theta_{t+1}) \quad (2)$$

We claim that this functional equation has a unique solution. To show this, let W be the set of all possible SPE reservation-wage functions. These are functions that associate a real number with every history (h_{t-1}, θ_t) . The reservation wage is equal to the outside option plus the discounted sum of future rents. Therefore, its value at every history is bounded by some finite number (as the maximal rent that the principal would pay at any period is less than 1). For every θ_t and function $w \in W$, define

$$q(w) \equiv \max_{\theta_{t+1}|\theta_t} w(\theta_{t+1}) - \mathbb{E}[w(\theta_{t+1}) \mid \theta_t]$$

For any pair $w, w' \in W$ define

$$d(w, w') \equiv |q(w) - q(w')| + \max_{\theta_{t+1}|\theta_t} |w(h_t, \theta_{t+1}) - w'(h_t, \theta_{t+1})|$$

It is straightforward to verify that d is a metric. Hence, (W, d) is a complete metric space.

Let $H(w)$ be a self-map on W defined by the R.H.S. of (2) (to economize on notation, we suppress the dependence of w on the history (h_{t-1}, θ_t)). This self-map is a contraction in (W, d) . To see this, note that for any pair $w, w' \in W$ and for any history (h_{t-1}, θ_t) ,

$$q(H(w)) = q(H(w')) = \max_{\theta_{t+1}|\theta_t} \theta_{t+1} - \mathbb{E}(\theta_{t+1} \mid \theta_t)$$

and

$$\max_{\theta_{t+1}|\theta_t} |H(w) - H(w')| = \delta |q(w) - q(w')|$$

It follows that

$$d(w, w') = \frac{d(H(w), H(w'))}{\delta} + \max_{\theta_{t+1}|\theta_t} |w(h_t, \theta_{t+1}) - w'(h_t, \theta_{t+1})|$$

Hence, for any $\delta < 1$, there exists $K < 1$ such that $d(H(w), H(w')) < Kd(w, w')$, implying that $H(w)$ is a contraction. Therefore, by the Banach Fixed Point Theorem, there exists a unique fixed point $\bar{w} = H(\bar{w})$, which means that the functional equation (2) has a unique solution. Plug the definitions of e and w as in the statement of the theorem into (1) and (2) to verify that they constitute a solution. Therefore, this must be the unique solution.

Step 4: The principal has a unique equilibrium payoff.

Proof: Consider first the lowest possible equilibrium principal's payoff. Then, there must be a Markovian equilibrium that sustains it (because players can credibly threaten to continue playing the same equilibrium after every deviation). Any Markovian equilibrium must be characterized by the reservation wage \bar{w} that is given by (2). As we saw in Step 3, there is a unique Markovian equilibrium, hence the lowest possible equilibrium principal's payoff is uniquely pinned down by it. Next, we obtain an upper bound on the principal's equilibrium payoffs. Consider an auxiliary principal-agent problem, in which the principal commits ex-ante to a wage strategy w as well as the agent's acceptance strategy a , subject to the participation constraint given by Step 1, and the constraint that for every h_{t-1} , $e_t(h_{t-1})$ is consistent with the players' strategies. Clearly, a solution to this problem attains a payoff weakly above any subgame perfect equilibrium payoff. However, since all continuation games that begin with the same θ_{t-1} are identical, there must be a Markovian solution to the problem. Since the Markovian solution is unique, the upper bound on the principal's equilibrium payoff coincides with the lower bound. It follows that there is a unique subgame perfect equilibrium payoff for the principal. ■

The unique equilibrium has several noteworthy properties. First, it is Markovian: players follow a simple rule that depends only on θ_t (in the agent's case) or θ_{t-1} (in the principal's case). Second, the agent always produces an output of 1 along the equilibrium path. Third, the equilibrium wage is rigid, or sluggish, in the sense that it is totally unresponsive to the current shock ε_t . The wage at t is a weighted average of the expected and maximal values of θ_t conditional on θ_{t-1} , where the weight on the latter is $1 - \delta$. Fourth, observe that if $\bar{\varepsilon}$ is sufficiently large and δ is sufficiently close to one, the agent's participation wage can take negative values. However, his actual

equilibrium wage is of course strictly positive. Finally, the agent earns an expected discounted rent of $\bar{\varepsilon}$, namely the difference between the maximal and expected values of ε . As F is subjected to a mean preserving spread, $\bar{\varepsilon}$ weakly increases, and thus the agent's rent goes up. The rent is independent of the discount factor: a higher δ simply means greater smoothing of the rent over time. Our model thus establishes a link between two phenomena: wage rigidity and efficiency wages, and it links them to the fundamentals $\delta, \bar{\varepsilon}$.

Comment: The role of the assumption that $\lambda \rightarrow 0$

The assumption that $\lambda > 0$ is crucial for equilibrium uniqueness. If $\lambda = 0$, it is possible to sustain equilibria in which the principal pays $w_t = e_t$, where e_t can take any value below 1 and above the highest participation wage that is possible given h_{t-1} . In this case, the agent's wage (lagged) expectations are self-sustaining: the principal does not wish to cut the wage below e_t because this would result in loss of output.

If λ were bounded away from 0, the equilibrium wage path would change as follows. First, the reference wage e_t would be strictly below the maximal participation wage that is possible given h_{t-1} . As a result, the wage at t would cease to be purely a function of θ_{t-1} : it would coincide with e_t at relatively low realizations of ε_t but it would coincide with the (higher) participation wage at relatively high realizations of ε_t . Second, the agent's equilibrium rent would be lower than in the $\lambda \rightarrow 0$ limit. Since our main objective in this note is to characterize the maximal rent that a reference-dependent agent can get in his long-run relationship with the principal, we do not provide a detailed characterization of this more general case.

References

- [1] Akerlof, George A. (1982): "Labor Contracts as Partial Gift Exchange," *Quarterly Journal of Economics*, 97, 543-569.
- [2] Akerlof, George A. and Janet L. Yellen (1990): "The Fair-Wage Effort Hypothesis and Unemployment," *Quarterly Journal of Economics* 195, 255-284.
- [3] Bewley, Truman F. (1999): *Why Wages Don't Fall during a Recession*, Harvard University Press.
- [8] Compte, Olivier, and Andrew Postlewaite (2004): "Confidence-Enhanced Performance," *American Economic Review*, 1536-1557.

- [8] Eliaz, Kfir, and Rani Spiegler (2013): "Reference-Dependence and Labor-Market Fluctuations," NBER Macroeconomics Annual 28.
- [8] Fang, Hanming, and Giuseppe Moscarini (2005): "Morale hazard," *Journal of Monetary Economics*, 52(4), 749-777.
- [8] Geanakoplos, John, Pearce, David and Ennio Stacchetti, (1989): "Psychological games and sequential rationality," *Games and Economic Behavior*, 1(1), 60-79.
- [8] Fehr, Ernst, Goette, Lorenz and Christian Zehnder (2009): "A Behavioral Account of the Labor Market: The Role of Fairness Concerns," *Annual Review of Economics*, 1, 355-384.
- [9] Kőszegi, Botond and Matthew Rabin (2006): "A Model of Reference-Dependent Preferences," *Quarterly Journal of Economics*, 121, 1133-1165.