Economics Letters 165 (2018) 6–9

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

# Managing intrinsic motivation in a long-run relationship☆

Kfir Eliaz [a,b], Ran Spiegler [a,c,d,*]

[a] Eitan Berglas School of Economics, Tel Aviv University, P.O.B. 39040, Ramat Aviv, Tel Aviv, 69978, Israel
[b] Department of Economics and Business Economics, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark
[c] Economics Department, University College London, Drayton House, 30 Gordon St, Kings Cross, London WC1H 0AX, United Kingdom
[d] Centre for Macroeconomics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom

## HIGHLIGHTS

- We study the effects of reference dependence on a repeated principal–agent interaction.
- The agent is productive as long as his wage does not fall below a "reference point".
- The reference point is the agent's lagged-expected wage.
- We characterize the game's unique Markov perfect equilibrium.
- The equilibrium exhibits wage rigidity. The agent's rent is equal to the maximal shock value.

## ARTICLE INFO

*Article history:*
Received 27 September 2017
Received in revised form 4 January 2018
Accepted 12 January 2018
Available online 31 January 2018

*Keywords:*
Reference-dependence
Dynamic contracting
Principal–agent
Wage rigidity
Intrinsic motivation

## 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 as long as his wage does not fall below a "reference point", which is defined as his lagged-expected wage in that period. We characterize the game's unique Markov 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, think of a parent hiring a nanny, or a hospital employing a surgeon.

Intrinsic motivation is a *dynamic* property — an agent who is initially motivated may temporarily *lose* his motivation in the course of his relationship with the principal. In addition, 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 that reflect changes in the external environment 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.

This paper studies a simple dynamic principal–agent model that explores this trade-off. The principal makes a "spot" wage offer at every period, and the agent decides whether to accept it. Once the agent rejects an offer, the two parties are permanently separated and the agent receives an outside payment $\theta_t$ at every $t$, where $\theta_t$ evolves according to some Markov process. The agent's output is reference-dependent, dropping from its normal level to zero whenever his wage drops below his reference wage $e_t$ by more than $\lambda > 0$.

Inspired by Kőszegi and Rabin (2006), we assume that the agent's reference wage is equal to the "rational" expectation of

☆ 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.

* Corresponding author.
*E-mail addresses:* kfire@post.tau.ac.il (K. Eliaz), rani@post.tau.ac.il (R. Spiegler).

https://doi.org/10.1016/j.econlet.2018.01.018
0165-1765/

his wage at period $t$ (conditional on continued employment), calculated at the end of period $t-1$ according to the parties' continuation strategies. The expectational aspect of the reference point 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. The lagged-expectations aspect captures the idea that like habits, reference points are sluggish in adapting to new circumstances.[1]

Our task is to characterize Markov perfect equilibria in this game, where the state at period $t$ is $(\theta_t, \theta_{t-1})$. To illustrate the possible effects of reference dependence, consider first the case of perfectly myopic parties. The agent's participation wage at period $t$ is $\theta_t$. Assume that $\theta_t$ can take two values, $\underline{\theta}$ and $\overline{\theta}$, with equal probability (independently of the history), where $\underline{\theta} < \overline{\theta} < 1$. Suppose that in equilibrium the parties' relationship is not severed at $t$ for any realization of $\theta_t$. Let $w(\theta)$ denote the equilibrium wage when $\theta_t = \theta$. Then, $e_t = \frac{1}{2}w(\underline{\theta}) + \frac{1}{2}w(\overline{\theta})$. If the principal paid the agent his participation wage in equilibrium, we would have $e_t = \frac{1}{2}\underline{\theta} + \frac{1}{2}\overline{\theta} > \underline{\theta}$. If $\lambda$ 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 $\lambda \to 0$ limit is $w(\underline{\theta}) = w(\overline{\theta}) = \overline{\theta}$.

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.

When parties are not myopic, 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 problem constitutes the paper's core. The unique solution to this problem extends the wage-rigidity effect of the myopic example: the equilibrium wage at any period is not responsive to the current shock, and the agent's discounted rent is the same as in a one-period model.

This note follows up Eliaz and Spiegler (2013), which essentially embedded an elaborate version of the myopic case in a search-matching model of the labor market.[2] The technical challenge in Eliaz and Spiegler (2013) arose from the possibility of rematching. Here we abstract from this complication and focus on the pure principal–agent relationship and the new considerations that arise from its infinite horizon. Re-incorporating it in a larger model of the labor market is a challenge for future research.

## 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 $t = 1, 2, \ldots$, 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 $\theta_s$ (0) at every period $s \geq t$. We assume that $\theta_t = \Psi(\theta_{t-1}) + \varepsilon_t$, where $\Psi$ is a deterministic function and $\varepsilon_t$ is *i.i.d* according to a *cdf* $F$ with mean zero. Let $\bar{\varepsilon}$ denote the highest value that $\varepsilon_t$ can take. We assume that $\Psi$ and $F$ are such that $\theta_t$ always takes values in $(0, 1)$.[3]

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 $\mathbf{1}(w_t \geq e_t - \lambda)$ as the agent's *output* in period $t$. The parameter $\lambda$ captures the tolerance of the agent's intrinsic motivation to frustrated wage expectations.[4] However, our analysis will focus on the $\lambda \to 0$ limit. 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, the principal's payoff 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 $\theta_t$. A strategy for the agent is a function $a$ that specifies for every $(h_{t-1}, \theta_t)$ 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 equal to the agent's lagged-expected wage at period $t$. More precisely, consider a history at the end of period $t-1$ (i.e., before $\theta_t$ 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 a conventional extensive game, but an extensive *psychological game* in the sense of Geanakoplos et al. (1989). However, since the belief-dependence is straightforward, we can work with the usual and familiar Subgame Perfect Equilibrium 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.

For simplicity, we restrict attention to SPE that are Markovian, where the state in period $t$ is $(\theta_{t-1}, \theta_t)$. Thus, a Markov Perfect Equilibrium (MPE) is a triple $(w, a, e)$ that satisfies the following properties for every $(\theta_{t-1}, \theta_t)$. First, given $(w, a, e)$, the wage $w(\theta_{t-1}, \theta_t)$ maximizes the principal's discounted sum of expected payoffs. Second, for every wage offer $w_t$, the decision $a(\theta_{t-1}, \theta_t, w_t)$ 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

$$e(\theta_{t-1}) = \mathbb{E}[w(\theta_{t-1}, \theta_t) \mid \theta_{t-1}\, ;\ a(\theta_{t-1}, \theta_t, w(\theta_{t-1}, \theta_t)) = 1]$$

and $e(\theta_{t-1}) = 0$ if the event $\{\theta_{t-1}, \theta_t \mid a(\theta_{t-1}, \theta_t, w(\theta_{t-1}, \theta_t)) = 1\}$ is null for the given $\theta_{t-1}$.

## 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 $\lambda = \infty$.)

[1] For earlier models in which an agent's productivity depends directly on his beliefs, see Compte and Postlewaite (2004) and Fang and Moscarini (2005).

[2] Effective myopia arose from a short horizon of the employment relation, rather than from a zero discount factor.

[3] E.g., $\Psi(\theta) = \alpha\theta + (1-\alpha)\frac{1}{2}$ and $F \sim U[-\bar{\varepsilon}, \bar{\varepsilon}]$, where $\bar{\varepsilon} \in (0, \frac{1}{2}(1-\alpha))$.

[4] Eliaz and Spiegler (2013) assume a stochastic, multiplicative version of reference-dependent output.