Signaling and motivation crowding out in gift exchange*

Daijiro Kawanaka[a]

October 10, 2025

Abstract

We theoretically study a simple principal-agent model called "gift-exchange" (Akerlof, 1982), in which a principal offers a fixed wage to an agent, and after observing the wage, the agent chooses his effort level. In our analysis, the agent is assumed to be guilt averse in the sense that he feels disutility when the realized outcome for the principal falls short of what he believes she expected from him. In this setting, the principal's positive wage offer works as a costly signal that she believes that the agent exerts a positive effort. According to this intuition, we show that positive wage and effort levels may arise in the equilibrium. Furthermore, we show that excessive wages may crowd out the agent's motivation.

Keywords: costly signaling, efficiency wage, psychological sequential equilibrium, motivation crowding out, self-fulfilling equilibrium

JEL Classification Numbers: C72, C91, D01

[*] We gratefully acknowledge financial support from Grant-in-Aid for JSPS Fellows from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and Waseda University Grant for Special Research Project. This paper has no conflicts of interest to disclose.

[a] School of Commerce, Waseda University. E-mail: kawanaka.daijiro@gmail.com

1. Introduction

In this study, we analyze a simple principal–agent framework often referred to as the "gift-exchange" model (Akerlof, 1982; Akerlof, 1984). A principal (she) offers a fixed wage to an agent (he), who, after observing the wage, chooses his effort level. Note that, unlike the standard contract theory literature, we assume no uncertainty and no information asymmetry in our one-shot model. Rather than relying on information frictions or repetition, our key assumption is guilt aversion: the agent experiences disutility when the realized outcome for the principal falls short of what he believes she expected from him. Under this assumption, the wage offer can serve as a costly signal of the principal's expectation that he will exert positive effort, and this expectation can be self-fulfilling. We obtain two main results. First, positive wages and effort levels can arise in equilibrium under guilt aversion; by contrast, in the benchmark case that maximizes material payoffs, both levels must be at their lowest feasible values. This is consistent with the efficiency-wage literature. Second, higher wages can crowd out the agent's motivation: increasing the wage does not necessarily induce greater effort and may even reduce it.

In the empirical literature, even a fixed wage can elicit higher effort (e.g., Gneezy and List, 2006; Bellemare and Shearer, 2009; Cohn et al., 2015). This type of prosocial behavior is often explained by models of reciprocity (e.g., Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006). However, there is also counterevidence to the gift-exchange hypothesis. For example, Gneezy and Rustichini (2000) document a nonmonotonic effect of monetary incentives on performance: introducing a small payment reduced performance relative to no payment, whereas within the positive-payment range higher pay increased performance—implying a discontinuity at zero pay ("pay enough or don't pay at all"). By contrast, other studies report an inverse U-shaped relationship between effort and incentive intensity—effort rises with incentives up to a point, after which stronger incentives reduce effort (Pokorny, 2008; Ariely et al., 2009; Esteves-Sorenson and Broce, 2022). Such phenomena—in which excessive incentives crowd out agents' intrinsic motivation—are commonly referred to as "motivation crowding out." There exist many theoretical accounts (e.g., Bénabou and Tirole, 2006), yet, to our knowledge, none explains all three patterns simultaneously. Our model reconciles these three seemingly contradictory findings.

This proceedings paper is organized as follows. Section 2 sets up the model. Section 3 analyzes the model. Section 4 concludes. In this proceedings paper, all proofs are abbreviated. In the full paper, we also examine the monotonicity of the agent's expectations and the uniqueness of the separating equilibrium.

2. Model

We consider a two-stage principal–agent model, referred to as a gift-exchange model (Akerlof, 1982). In our model, a principal (she) offers a wage $w \in \mathbb{R}_+$ to an agent and maximizes her payoff

$$pe - w, \tag{1}$$

where $p \in \mathbb{R}_{++}$ denotes productivity, demand, or the output price. Observing the wage level w, the agent (he) chooses whether to accept the offer; if he accepts, he chooses an effort level $e \in \mathbb{R}_+$. We assume no uncertainty, no information asymmetry, and no repetition, unlike the literature of the standard contract theory. Note that the wage is fixed, i.e., it cannot be conditioned on effort or output. The agent's payoff when he accepts the offer w and chooses e is assumed to be

$$U(w,e) = w - c(e) - g(r-e), \tag{2}$$

where a differentiable function $c : \mathbb{R}_+ \to \mathbb{R}_+$ denotes the agent's cost, a differentiable function $g : \mathbb{R} \to \mathbb{R}_+$ denotes his guilt, and r denotes his second-order belief about what effort level she believes that he chooses. Assume $g'(x) > 0$ for all $x > 0$ and $g(x) = 0$ for all $x \leq 0$. For example, if $g''(x) > 0$ holds, then we can interpret that the marginal guilt is decreasing, i.e., $\frac{\partial^2 g(r-e)}{\partial e^2} < 0$ for any $e < r$. Jensen and Kozlovskaya (2016) caracterize a specific form of such a representation in which a marginal guilt is decreasing. When the marginal guilt is decreasing, it is an important assumption that $c'(0) \leq g'(0+)$ for Main result 1. However, our analysis permits that $g''(x) < 0$ to explain motivation crowding out (Main result 2). We say that the agent is guilt-neutral if the guilt function is constant, and we analyze the guilt-neutral case as a benchmark in our analysis. If he does not accept the offer, each payoff is normalized to zero.

Battigalli and Dufwenberg (2009) generalize Kreps and Wilson's (1982) sequential equilibrium to formulate "psychological sequential equilibrium." Following their framework, we define an equilibrium concept as follows:

Definition (Psychological Sequential Equilibrium). An assessment (σ^*, r^*) is a *Psychological Sequential Equilibrium* (PSE) if it satisfies

- **Consistency.** There exists a sequence of fully mixed strategy profiles $\{(\sigma_P^k, \sigma_A^k)\}_{k\geq 1}$ such that $\sigma_P^k \to \sigma_P$ and $\sigma_A^k(\cdot \mid w) \to \sigma_A(\cdot \mid w)$ for each $w \in \mathbb{R}_+$, and

$$r(w) = \lim_{k\to\infty} \mathbb{E}_{e\sim\sigma_A^k(\cdot|w)}[e] \tag{3}$$

for each $w \in \mathbb{R}_+$.

- **Sequential rationality.** (i) For every w, $\mathrm{supp}(\sigma_A(\cdot \mid w)) \subseteq \arg\max_{e\geq 0}\{w - c(e) - g(r(w) - e)\}$; (ii) At the initial node, $\mathrm{supp}(\sigma_P) \subseteq \arg\max_{e\geq 0}\{p\mathbb{E}_{e\sim\sigma_A(\cdot|w)}[e] - w\}$.

In words, the first condition says that the both players have "right" beliefs in a sense that the agent's second-order belief is identical to the principal's first-order belie, and her belief is identical to his actual choice. The second condition says that each player maximizes each utility with respect to each belief, permitting that the agent exhibits guilt aversion.

3. Main results

In the benchmark case in which the agent feels no guilt, he exerts the minimum effort for any wage, and thus, the principal offers the minimum wage in the subgame perfect equilibrium:

Benchmark result. If a guilt function g is constant, then there is a unique PSE in which $w^* = e^* = 0$.

In contrast, if the agent is guilt averse, gift exchange may arise in equilibrium. The guilt averse agent chooses $e^* = r^*$ on the equilibrium path, and if w works as a costly signal of r, a higher level of wage induces a higher level of effort. This fact is summarized as follows:

Main result 1. There exists a PSE in which w^* and e^* are positive if and only if there is $r > 0$ such that

$$c'(r) \leq g'(0+) \tag{4}$$

and

$$pr - c(r) \geq 0. \tag{5}$$

Note that if $g''(0) > 0$, then $c'(0) \leq g'(0+)$ implies conditions (4) and (5) for $r = 0$. In particular, if $g''(x) > 0$ for each $x \geq 0$, there is a single crossing point $r^* > 0$ in which $c(r^*) = g(r^*)$ and the agent chooses $e^* = r^*$ in PSE. If g'' is non-monotone, then the effort level $e^*(r)$ becomes non-monotone. That is, too much expectation may crowd out his motivation. This fact is summarized as follows:

Main result 2. The agent's effort $e^*(r)$ is strictly increasing in r if and only if $g''(r - e) > 0$; it is strictly decreasing if and only if $g''(r - e) < 0$.

There exist many theoretical papers which imply that too much incentive crowds out agents' motivations (e.g., Holmström and Milgrom, 1991; Bénabou and Tirole, 2003; Bénabou and Tirole, 2006; Ellingsen and Johannesson, 2008), but our result suggests another mechanism from them. We can suppose an S-shaped guilt function, in which $g''(x) < 0$ in some area of x and $g''(x) > 0$ in another area of x. Then, as the wage–induced belief $r = \mu(w)$ moves across regions where g'' switches sign, the comparative statics of the agent's best response also switch. Let $e^*(r)$ solve the interior FOC $c'(e) = g'(r - e)$ with the complementary slackness condition that $e^*(r) = 0$ whenever $e \geq r$ (so $g(r - e) = 0$). By the implicit function theorem, whenever $e^*(r) \in (0, r)$,

$$\frac{de^*(r)}{dr} = \frac{g''(r - e^*(r))}{c''(e^*(r)) + g''(r - e^*(r))}. \tag{6}$$

Hence the sign of $\frac{de^*}{dr}$ is the sign of g'' evaluated at the equilibrium deviation $r - e^*(r)$.

4. Conclusion

Assuming that the agent has guilt-averse preferences, this proceedings paper shows that (1) there exists a psychological sequential equilibrium in which gift exchange arises, and (2) motivation crowding out can occur. In addition, we examine the monotonicity of the agent's expectations and the uniqueness of the separating equilibrium in the full paper.

References

[1] Akerlof, G. A., 1982. Labor contracts as partial gift exchange. The Quarterly Journal of Economics, 97(4), 543-569.

[2] Akerlof, G. A., 1984. Gift exchange and efficiency-wage theory: Four views. The American Economic Review, 74(2), 79-83.

[3] Ariely, D., Gneezy, U., Loewenstein, G., and Mazar, N. 2009. Large stakes and big mistakes. The Review of Economic Studies, 76(2), 451-469.

[4] Battigalli, P., and Dufwenberg, M., 2009. Dynamic psychological games. Journal of Economic Theory, 144(1), 1-35.

[5] Bellemare, C., and Shearer, B., 2009. Gift Giving and Worker Productivity: Evidence from a firm-level experiment. Games and Economic Behavior, 67(1), 233–244.

[6] Bénabou, R., and Tirole, J., 2003. Intrinsic and extrinsic motivation. The Review of Economic Studies, 70(3), 489-520.

[7] Bénabou, R., and Tirole, J., 2006. Incentives and prosocial behavior. American Economic Review, 96(5), 1652-1678.

[8] Cohn, A., Fehr, E., and Goette, L., 2015. Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment. Management Science, 61(8), 1777-1794.

[9] Dufwenberg, M., and Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games and Economic Behavior, 47(2), 268-298.

[10] Ellingsen, T., and Johannesson, M., 2008. Pride and prejudice: The human side of incentive theory. American Economic Review, 98(3), 990-1008.

[11] Esteves-Sorenson, C., and Broce, R., 2022. Do monetary incentives undermine performance on intrinsically enjoyable tasks? A field test. Review of Economics and Statistics, 104(1), 67-84.

[12] Falk, A., and Fischbacher, U., 2006. A theory of reciprocity. Games and Economic Behavior, 54(2), 293-315.

[13] Gneezy, U., and List, J. A., 2006. Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. Econometrica, 74(5), 1365–1384.

[14] Gneezy, U., and Rustichini, A., 2000. Pay enough or don't pay at all. The Quarterly Journal of Economics, 115(3), 791-810.

[15] Holmström, B., and Milgrom, P., 1991. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. The Journal of Law, Economics, and Organization, 7, 24-52.

[16] Jensen, M. K., and Kozlovskaya, M., 2016. A representation theorem for guilt aversion. Journal of Economic Behavior & Organization, 125, 148-161.

[17] Kreps, D. M., and Wilson, R., 1982. Sequential equilibria. Econometrica, 50(4), 863-894.

[18] Pokorny, K., 2008. Pay—but do not pay too much: An experimental study on the impact of incentives. Journal of Economic Behavior & Organization, 66(2), 251-264.