Why Do External Rewards Crowd Out Intrinsic Motivation While Self-Rewards Do Not?

Ksenia Panidi

November 26, 2013

Abstract

The phenomena of crowding-out and crowding-in of motivation after an externally administered reward have received empirical support in the domains of health care and education. In addition, self-rewards have been observed to provide motivational crowding-in but not crowding-out. In the present paper a theoretical model explaining the observed differences in the effects of external and self-rewards is developed. The model is based on the combination of the dual-self approach to the analysis of the time-inconsistency problem with the principal-agent framework. It is shown that psychological property of disappointment aversion may help to explain these differences in the situation when abstention costs are not perfectly known in advance.

JEL Classification: D80

Keywords: self-control, intrinsic motivation, self-rewards, motivational crowding-out

Introduction

Recent behavioral economic research has drawn our attention to the problem of time-inconsistent behavior, or, in other words, the self-control problem. An individual demonstrating such behavior is characterized by present-biased preferences, according to which current consumption has a disproportionately higher weight compared to future consumption. This phenomenon was suggested to be underlying such problems as various types of addiction (like smoking and alcoholism), failures to keep diet, poor sugar level control by diabetics and procrastination in performing onerous tasks (see Kan (2007), Dodd (2008), O'Donoghue and Rabin (1999)).

One of the popular methods of dealing with the self-control problem consists in providing an individual with a reward contingent on successful accomplishment of a task. Rewards can be offered in a a tangible form (such as money or a possibility to win a prize), or in a non-tangible form (such as praise, encouragement or verbal approval). In the present chapter I will distinguish between externally and internally provided rewards (or self-rewards). External rewards are usually offered by people interested in promoting prudent behavior in those individuals whom they care about. Common examples of such situations are payment of money by parents to their children for good grades, encouragement by one of the spouses of the other to quit smoking, or participation in a program rewarding a successful weight loss over a certain period of time. Self-rewards are those that a person offers him/herself to overcome the problem of time-inconsistency. A common example is rewarding oneself with a gift for every achievement of a goal (like abstention from smoking for one day or meeting a deadline for writing a report).

The aim of the present chapter is to theoretically investigate and compare the short-run and long-run effects of rewards (both external and internal) as a means to encourage self-control. The need for such analysis stems from the ongoing debate on the efficiency of externally provided rewards to increase the level of self-control efforts.

Empirical evidence on the effects of rewards of both types is mixed. A wide range of studies suggests that external rewards may help to promote self-control behavior in the short-run at the cost of its poor long-term maintenance. Jochelson (2007) reviews 41 studies on the effects of monetary rewards observed in various health-improving programs. These programs offer various kinds of rewards (direct money payments, gift certificates, lottery prizes) for participation and/or achievement of certain health-related goals. Some studies mentioned in this review indicate that the use of rewards increases the rates of participation in the program (see Bains et al. (1998), Englberger (1999), Harland et al. (1999), Hey and Perera (2005)). Positive effects were observed in the intervention period in smoking-cessation and cocaine abstinence programs. However, these studies document significant relapse rates and poor long-term maintenance of the results achieved during participation.

Several studies offer a more optimistic view on the use of external rewards (see Volpp et al. (2008), Volpp et al. (2009), Finkelstein et al. (2007)). In these experiments participants were rewarded with money or lottery prizes for their achievement in weight-loss or smoking-cessation program. The maintenance of these achievements was monitored several months after the intervention period. The results of these studies indicate that for some people external rewards may generate a larger crowding-in of motivation (as measured by the share of the participants who maintained the achievement after the intervention period) compared to the control group.

Little is known about the factors that might determine whether an external reward will lead to crowding-in or crowding-out of motivation for a particular individual. An interesting experiment of Leuven et al. (2006) suggests that ability to succeed in an activity that is being rewarded might be such a factor. In this study students with high abilities in math have demonstrated a boost in motivation during the next three years after the payment year. By contrast, students with low abilities have shown significantly lower results during the payment year which have reduced even further in three subsequent years.

Empirical evidence on the effects of self-rewards suggests that they may outperform externally provided incentives both in the short- and in the long-run. The results of the experimental study of Ryan et al. (1995) show that patients subject to alcohol treatment demonstrated greater involvement and better treatment retention when their motivation was mostly internal. Williams et al. (1996) and Williams et al. (2002) document that better short-term and long-term results of participation in a weight-loss and smoking-cessation programs were associated with more autonomous (as opposed to more controlling) motivation. Curry et al. (1991) finds similar pattern for the level of intrinsic *relative* to extrinsic motivation in smoking cessation.

Although several theoretical papers exist that attempt to explain the phenomenon of motivational crowding-out and crowding-in using economic analysis (Bénabou and Tirole(2003), Harvey (2005)), the direct comparison between external and self-rewards is underexplored.

The theoretical model developed in this chapter demonstrates that crowding-in and crowding-out of motivation may be observed under both types of rewards if the self-control costs are not perfectly known. However, which effect will take place depends on the level of the self-control costs and on the individual confidence in the ability to abstain. Moreover, both effects are generated under the assumption that an individual is disappointment-averse. The individual dislikes learning that her self-control costs are high, since this is discouraging for her future attempts at self-control.

The model is based on the dual-self approach to the problem of self-control. I follow the modeling practice for this approach, introduced in Fudenberg and Levine (2006), according to which a decision of the individual to consume a tempting but harmful product could be represented as an equilibrium in a game between a multi-period forward-looking self and a series of short-run one-period agents. Each of the short-run selves maximizing its utility prefers to consume as much of the product as possible. The long-run self understands that such an unbounded consumption can have negative long-run consequences and, hence, tries to restrict it. In Fudenberg and Levine (2006) the long-run self does this by making sure that each short-run self simply does not have an access to large amounts of a product that could be overconsumed. In my model, rather than restricting access to resources, the long-run self (or the self-principal) incentivises each of the short-run selves (or agents) to consume less by paying a bonus dependent on the abstention level. To model externally administered rewards I consider the situation where in the first period an external principal pays a bonus in addition to that of the self-principal. In the second period external intervention is removed and the individual has to decide on the level of self-reward. The representation of the dual-self model in the "principal-agent" framework allows for a direct comparison between the short-term and long-term effects of external and self-rewards.

There are two important ingredients in the model. One is the assumption that abstention costs are initially not perfectly known but can be revealed in case the agent exerts high effort in the first period. This seems to be a natural assumption: for example, if an individual has little experience in cessation of smoking, she may not know how hard it is to quit smoking. But once she tried to quit, she gets more precise perception of the costs of abstention. Second ingredient of the model is the assumption that self-principal is disappointment-averse. I model disappointment aversion as loss aversion with respect to the expected value of a lottery. In my model an individual may get disappointed if after the first period high costs of abstention have been revealed. Knowing that abstention costs are high leads to loss of motivation in the second period. Therefore, revealing those costs may not be desirable from the perspective of the first period, especially if an individual expects her costs to be high with probability close to 1. This means that in order to avoid disappointment and preserve self-confidence in one's ability to quit smoking, an individual may not motivate herself to work too hard.

On the contrary, an external motivator may not be worried about the self-confidence of the agent after the intervention period and may not be subject to disappointment aversion[1]. She therefore may induce higher efforts compared to those induced by an individual herself, which results in the loss of motivation for those agents who tried to abstain and observed high abstention costs.

Depending on the degree of loss aversion, the level of agent's self-confidence and his sensitivity to an external reward, the model predicts that three situations are possible. First, the presence of the external principal may not change the behavior of a person compared to that without an external reward. Second, external reward may make abstention more likely both in the short-run and in the long-run compared to "no intervention" condition. Finally, under the external reward some people may be more likely to abstain in the short-run, but less likely so in the long-run.

The Model

General setting

I consider a model where an individual lives for two periods. I follow the dual-self approach to the self-control problem by representing an individual as consisting of two selves in each period. The short-run self corresponds to the individual's temptation to consume a harmful good (e.g. cigarettes) with no concern regarding the future negative consequences of this consumption. In each period there is a new short-run self and each of these selves is fully myopic, i.e. concerned only about a one-period

[1] The issue of how much an external motivator takes into account future effects of her actions depends to a large extent on the context of the situation. We believe that there exist many situations in which external principal may not take into account the long-run consequences of her actions. For example, the role of an external motivator may be played by a company like Tangerine Wellness or Weight Wins that offer rewards for achieving some self-control goal. These companies may focus only on positive short-run effects of rewards for advertisement purposes. However, even if an external motivator has altruistic intentions towards the agent (e.g. a parent or a teacher), she may still demonstrate a high degree of myopia with respect to the future consequences of an excessive control (Soenens et al.(2007), Darling (1999)).

utility. The long-run self corresponds to the "prudent" part of an individual willing to promote abstention from a harmful product. This self is forward-looking and is assumed to gain utility from abstention in each period.

As we are exploring the effect of externally and self-administered rewards it is convenient to adhere to the principal-agent framework in modeling the relationship between short-run and long-run self. Therefore, each of the short-run selves will be called an "agent", while the long-run self will play the role of the "*self*-principal".

A. Actions

Reward paid to the agent can be seen as a positive bonus offered by the principal to stimulate abstention. In each of the periods the self-principal decides on the level of bonus $b^{SP} \geq 0$ to offer to the agent.

In order to model externally administered rewards I assume that in the first period external principal offers a bonus $b^{EP} \geq 0$ in addition to that set by the self-principal. There are two possible cases I investigate. In the first one the external principal exists for only one period and cares only about the current (short-run) level of abstention, while in the second one the external principal exists for two periods, cares about outcomes in both periods and pays the bonus only in the first period.

In each period agents 1 and 2 decide on the level of effort to exert given the bonuses provided by either the self-principal alone or together with the external principal. An agent can choose effort level e_t, $t \in \{1,2\}$ from the set $\{0,1\}$. Table 1 summarizes the choice variables for both principals and the agents.

	Period 1	**Period 2**
Self-principal	First-period bonus $b_1^{SP} \geq 0$	Second-period bonus $b_2^{SP} \geq 0$
External principal	Bonus $b^{EP} \geq 0$	-
Agent 1	Effort level $e_1 \in \{0,1\}$	-
Agent 2	-	Effort level $e_2 \in \{0,1\}$

Table 1.

B. Types

The cost of zero effort is zero. Exerting effort of 1 is costly and the cost is equal to some $k > 0$. I assume that the cost of effort in the first period is not perfectly known and that an individual can be of two types: "strong" type is characterized by low cost of efforts $k = k_L$, while "weak" type has high abstention costs $k = k_H > k_L$. Without loss of generality one can set $k_L = 0$. I also impose an assumption that $k_H > 1$. As will be seen later, this assumption ensures that the abstention cost of a "weak" type is so high that in the second period the individual gets discouraged from abstention if the "weak" type is revealed. The type of the individual is determined by nature prior to the beginning of the first period and does not change between first and second period.

C. Beliefs

In the first period the type is not known, however both principals have some prior regarding the type of the agents. I assume that the priors of both principals coincide and denote the probability of types $\{k_L, k_H\}$ as $(1 - p)$ and p respectively. Since each agent and the self-principal represent different subdivisions of the same individual, one may logically assume that they share the same information and therefore have the same priors about the cost type. Therefore, the priors of the agent coincide with those of the self-principal.

After the first period an individual may receive new information about the costs of abstention. It seems natural to assume that the more effort the individual exerts, the more information she gets about the costs of these efforts. I make an assumption that if an agent chooses zero effort in the first period, then no new information is revealed about the costs and in the second period the self-principal's beliefs do not change. If in the first period an agent chooses $e_1 = 1$, then he observes his payoff and, hence, derives his true cost type. Hence, the second-period beliefs (of both the agent and the self-principal) regarding the cost type can be either equal to the first-period beliefs (in case of no revealing) or equal to the true type (in case of revealing).

D. Preferences

In order to specify preferences of agents and principals one needs to make assumptions regarding the utility that the agent gets from external and self-bonuses. In case when both external and self-principal are present agent 1 receives bonuses from both of them if he generates positive level of abstention. However, we assume that the agent may react differently to external or self-bonus. Recent psychological findings (namely, the Personality Systems Interaction Theory, or PSI) indicate that individuals may be described as either "state-oriented" or "action-oriented". State-oriented individuals are mostly motivated by external factors or goals (such as encouragement of parents and teachers or controlling framing of instructions). Action-oriented individuals, on the other hand, do not rely on the external sources of regulation as they are able to internally generate motivation for exerting efforts (see Baumann and Kuhl (2005)). Therefore, I assume that action-oriented individuals are relatively more sensitive to internal rather than external rewards when both are present, while it is vice versa for state-oriented individuals. To model this assumption the agent 1's utility function is presented in the following way:

$$U_1^A = b_1^{SP} + \gamma b^{EP} - k \cdot e_1. \tag{1}$$

Here parameter $k \in \{k_L, k_H\}$ denotes the cost of abstention, $e_1 \in \{0,1\}$ represents effort level, and $\gamma \geq 0$ indicates the relative sensitivity of the agent towards the external reward. In the second period external principal is no longer present, hence, agent 2's utility is simply:

$$U_2^A = b_2^{SP} - k \cdot e_2. \tag{2}$$

I now specify the utility functions of the principals. If the external principal cares only about the outcome of the first period, her utility function looks as follows:

$$U^{EP} = (1 - b^{EP}) \cdot e_1. \tag{3}$$

If the external principal cares about the outcomes in both periods, her utility function looks as follows:

$$U^{EP} = (1 - b^{EP}) \cdot e_1 + e_2. \tag{4}$$

In this chapter I first consider a benchmark case where the external principal is myopic and takes into account only the outcome of the first period. I then follow with its extension assuming that the external principal is forward-looking. She cares about the current abstention of the agent as well as about the future one, but is only able to provide a bonus in period 1. This case may correspond to many real-life situations. For example, a governmental program for quitting smoking may provide a bonus conditional on abstention during some limited period of time. However, this program may aim at promoting the long-term abstention, i.e. beyond the payment period. The same may be true, for instance, in the experimental settings in which students are being paid for good grades. Although the payment period is limited, the goal of the program might be to make students willing to exert more effort in the future as well.

Consider the self-principal's utility function. Denote by x_1 and x_2 the outcomes that the self-principal receives in the first and second period respectively, and by Ex_2 the expectation of the second-period outcome in the beginning of period 1. The outcome that she receives in the first period is either equal to zero (in case the agent does not abstain at all) or to $(1-b_1^{SP})$, where b_1^{SP} is the bonus paid to the agent for abstention. In the first period the self-principal does not know the agent's type exactly, but knows that it will be revealed if the agent chooses $e_1 = 1$. As a result her choice to induce a positive effort level in the first period implies that she simultaneously forms some expectation about the outcome of the second period. The main assumption here is that the self-principal is disappointment-averse, or in other words, loss-averse with respect to the expected outcome. Formally this means that overall the self-principal's ex-post utility from the perspective of period 1 when she had an expectation of Ex_2 consists of two parts: linear, reflecting the utility from abstention as such, and reference-dependent emotional part ($v(x_2)$), or an emotional response to the changes in the second-period outcomes. Formally, this utility function can be written in the following way:

$$U^{SP} = x_1 + E[x_2 + v(x_2)], \tag{5}$$

$$x_1 = (1-b_1^{SP})\cdot e_1,$$

$$x_2 = (1-b_2^{SP})\cdot e_2,$$

$$v(x_2) = \begin{cases} x_2 - Ex_2, & \text{if } x_2 \geq Ex_2; \\ \lambda(x_2 - Ex_2) & \text{if } x_2 < Ex_2, \end{cases}$$

with $\lambda > 1$ being the loss-aversion coefficient. The first term of the expression represents the outcome of the first period (x_1), the second one is the expected utility of the second period. This expression takes into account not only the present need to stimulate abstention but also its consequences for the self-confidence in the second period. If agent 1 has made an effort to abstain and revealed that he has low abstention costs, the self-principal realizes that in the second period abstention will be an easy task and will not require a large bonus. On the contrary, if agent 1 reveals that abstention costs are high, the self-principal's future profit lowers relative to its expectation (since she realizes that only a large bonus will suffice in period 2 for abstention). Because losses generally loom larger than gains and the self-principal is loss averse by assumption observing the agent to be a weak type leads to a larger absolute loss in utility than a gain produced by observing a strong type. This property is reflected by the loss-aversion coefficient λ. This two-part utility function reflects the trade-off that the self-principal is confronted with. On the one hand, she prefers to receive a higher physical payoff. On the other hand, she may have expectation of the outcome that she could have received and she, therefore, dislikes to fall short of this expectation. The self-principal gets additional positive utility when the outcome exceeds the expectation. Note that since external principal cares only about the outcome of the first period, I do not include disappointment-aversion in her utility function. In the present formulation neither the agent, nor the self-principal derive utility directly from the ability to abstain. Rather they derive utility from the outcome that having low abstention costs allows to achieve. Observing a high abstention cost is disappointing not by itself but because it lowers the expected outcome of the second period.

Equilibrium analysis

First, I analyze the equilibrium of a two-period game in which external principal is not present. The self-principal and both agents are characterized by the assumptions described above. Consider the agent's problem. Note that it is the same in both periods and may differ only in whether the agent knows her self-control costs or can only rely on his best guess which is the expected cost pk_H. In the first period the agent maximizes his expected utility given by the expectation of (1) with $b_{EP} = 0$, since there is no external principal. The agent does not know his costs of abstention, therefore, his expected utility is written as:

$$EU_1^A = b_1^{SP} - \hat{k}, \tag{6}$$

with $\hat{k} = pk_H$. In his first-period optimum the agent chooses:

$$e_1 = \begin{cases} 0, & \text{if } b_1^{SP} < \hat{k}; \\ 1, & \text{if } b_1^{SP} \geq \hat{k}, \end{cases} \tag{7}$$

In the second period the agent's problem solution is the same in case the true cost of abstention has not been revealed in period 1. If the agent has chosen $e_1 = 1$ then in the second period $\hat{k}$ will be substituted by the actual abstention cost.

Since the self-principal lives for two periods we solve her problem starting with period 2. The optimal bonus will depend on two factors: whether the true cost has been revealed and (in case of non-revealing) on the level of self-confidence p. Since $k_H > 1$ by assumption, then if the weak type has been revealed the agent will require a bonus larger than 1 to abstain in period 2. Obviously, the self-principal will not offer such bonus since her maximum outcome in case of abstention equals 1. Therefore, the second-period agent chooses $e_2 = 0$ and the self-principal's optimum is $b_2^{SP} = 0$. If the agent's cost of abstention is low, then he does not require any bonus to exercise self-control. Hence, $e_2 = 1$ and $b_2^{SP} = 0$ when $k = k_L$. However, when the abstention cost is not known, in equilibrium the self-principal chooses to incentivise the agent only when her self-confidence is high enough (i.e. when $p < \bar{p} = 1/k_H$). In the opposite case the agent would require too large bonus (at least equal to $pk_H > 1$) which would lead the self-principal with a negative second-period utility.

In order to determine the equilibrium of the first period we need to compare the expected utility (i.e. expectation of 5) that the self-principal obtains by paying the minimal bonus necessary for abstention and by not paying it.

Suppose the agent has chosen zero effort in period 1. When the self-confidence is high enough (i.e. the probability of high abstention costs is low, or $p \in [0, \bar{p}]$), in the second period the self-principal chooses to pay a bonus to the agent who then exerts effort $e_2 = 1$. In equilibrium this bonus will be equal to $\hat{k}$ and, hence, the self-principal's second-period utility will be $U_2^{SP}(e_1 = 0) = 1 - b_1^{SP} = 1 - \hat{k}$. In the opposite case, i.e. when $p \in (\bar{p}, 1]$ the self-principal cannot afford paying a sufficiently high bonus and receives utility of zero.

Suppose now that the agent has chosen $e_1 = 1$. Since the cost of abstention is not known in period 1 but is revealed in this case by the beginning of period 2, the self-principal does not know her exact second-period utility and may only compute its expectation. If the agent is of weak type then the self-principal may only get zero, while with a strong type her utility equals 1. Since the probabilities of these types are p and $(1 - p)$ respectively, the self-principal's expectation of utility (or her reference-point) is $(1 - p)$. After the agent's type is revealed and the self-principal knows exactly the utility of the second period, she experiences it either as a gain or as a loss relative to the reference point. This means that her first-period expected utility may be computed as a sum of two components according to (5) in the following way[2]:

$$EU_1^{SP}(e_1 = 1) = (1 - \hat{k}) + (-\lambda p(1 - p) + (1 - p)(1 + p)), \tag{8}$$

Proposition 1. *There exists $p^* \in (0, 1)$ such that in equilibrium:*

(1) The self-principal offers a bonus $b_1^{SP} = pk_H$ for any $p \leq p^*$ and pays zero for any $p > p^*$.

[2] See Appendix for the computation.

(2) For $1 < \lambda \leq k_H + 1$, $p^* = \frac{\lambda + k_H - \sqrt{(\lambda + k_H)^2 - 8(\lambda - 1)}}{2(\lambda - 1)} > \frac{1}{k_H}$.

(3) For $\lambda > k_H + 1$, $p^* = \frac{1}{\lambda - 1} < \frac{1}{k_H}$.

(4) The payment threshold p^* *decreases in* λ.

The second proposition describes the equilibrium choice of efforts in the first and second period given the bonus paid by the self-principal according to Proposition 1. Since the agent's type is not known, after the agent has chosen $e_1 = 1$, we may only compute the expected second-period level of efforts which will depend on the belief p. Obviously, if no revealing has taken place, the second-period efforts will not depend on p in equilibrium. I denote this expected effort level by $\bar{e}_2$. The following proposition states the equilibrium effort level in the first and second period depending on loss aversion and belief p.

Proposition 2.

(i) *For* $1 < \lambda \leq k_H + 1$*:*

if $p \in [0, p^*]$, *then* $b_1^{SP} = pk_H$, $b_2^{SP} = pk_H$, $e_1 = 1$, $\bar{e}_2 = 1 - p$;

if $p \in (p^*, 1]$, *then* $b_1^{SP} = 0$, $b_2^{SP} = 0$, $e_1 = 0$, $\bar{e}_2 = 0$.

(ii) *For* $\lambda > k_H + 1$*:*

if $p \in [0, p^*]$, *then* $b_1^{SP} = pk_H$, $b_2^{SP} = 0$, $e_1 = 1$, $\bar{e}_2 = 1 - p$;

if $p \in (p^*, 1/k_H]$, *then* $b_1^{SP} = 0$, $b_2^{SP} = pk_H$, $e_1 = 0$, $\bar{e}_2 = 1$;

if $p \in (1/k_H, 1]$, *then* $b_1^{SP} = 0$, $b_2^{SP} = 0$, $e_1 = 0$, $\bar{e}_2 = 0$.

The intuition behind these propositions is straightforward. The self-principal faces a trade-off. On the one hand, she may incentivize the provision of effort in the first period, which gives her an immediate payoff. However, exerting effort also reveals the agent's type. If high cost of effort is revealed, this has two effects on the self-principal's payoff. First, it means that in the second period the agent will be discouraged to exercise self-control. He will require a bonus too high for the self-principal to pay. Second, since the self-principal is loss averse with respect to her expected payoff, she experiences disappointment from learning that her second-period payoff is lower than she might have expected to obtain prior to self-control cost being revealed. Alternatively, the self-principal may decide not to stimulate abstention in the first period. In this case the agent's type is not revealed and the self-principal does not experience any disappointment. However, whether the agent will exert effort in the second period or not will depend on his belief p (or his level of self-confidence $(1-p)$): whenever $p < 1/k_H$ the agent will exert effort. Therefore, the self-principal's trade-off consists in the choice between abstention in the first period at the expense of a potential loss of motivation in the second period, and foregoing first-period benefits of abstention to maintain self-confidence in the future.

The second proposition describes the conditions under which crowding-in and crowding-out of motivation may be observed. Obviously, individuals with very low probability of being a strong type do not make efforts to abstain in any period. In this case the agent always requires a bonus higher than what the self-principal may afford to pay. In particular, this happens for any individual with $p > max\{p^*, 1/k_H\}$. The situation is different when the self-confidence is sufficiently high. If loss aversion λ is smaller than the threshold (k_H+1), the self-principal is not very disappointment-averse as she does not experience a big loss from observing a weak type. Hence, she will pay a bonus for any $p \in [0, p^*]$. Exerting effort in the first period reveals the self-control cost, and only agents with low cost k_L will decide to abstain in the second period. In other words, when loss aversion is low, the self-reward leads to motivational crowding-out for those with high abstention costs. For those with low costs the motivation for self-control is increased since in the second period they do not require any bonus to exert effort. Similar situation is observed when loss aversion is high $(\lambda > k_H + 1)$. The higher is loss aversion, the less the self-principal is willing to pay a bonus because of the fear to be disappointed by high abstention costs. When loss aversion is high enough, the payment threshold p^* is smaller than the minimal self-confidence needed for the agent to exert effort without any bonus. In this case the agent with an intermediate self-confidence ($p^* < p < 1/k_H$) will choose zero effort in the first period, but his self-confidence will be enough to exercise self-control in the second period.

In order to determine the effect of external rewards on motivation I assume that the external principal offers a bonus together with the self-principal in the first period. External principal either exists and gains utility only in period 1, or exists and gains utility in both periods. The agent's utility function is determined by (1). The self-principal solves her problem for every level of external reward considering it as given. Knowing the solution of the choice of the self-principal external principal decides on the optimal level of reward to offer.

I start the analysis assuming that the external principal exists only for one period.

Solving the model in this case establishes the conditions for two different outcomes of the external principal's participation. Namely, when the external principal is present two different situations are possible. In the first one, the self-principal does not incentivize abstention in the first period, yet the agent's self-confidence is high enough ($p \in [p^*, 1/k_H]$) that the agent chooses the effort of 1 in the second period. Note that the agent's self-confidence is maintained because the self-control cost has not been revealed. If the external principal is present in one period only, she does not care about the agent's future self-confidence, and may choose to provide a bonus in period 1 to get immediate benefits of abstention. This may crowd out the second-period motivation for those agents who discover high abstention costs. In other words, for the same level of self-confidence, in the presence of an external reward the agent may be more likely to abstain in the short-run, but less likely to do it in the long-run (compared to the case of the self-principal alone).

The second important case is one in which the agent's self-confidence is so low that in the absence of external bonus he chooses zero effort in both periods ($p > 1/k_H$). Then external principal may incentivize the agent to abstain in period 1. With probability

$(1-p)$ low abstention cost is revealed, and, hence, with probability $(1-p)$ the agent abstains in period 2. This means that for a sufficiently low self-confidence, the agent is more likely to abstain in both periods under the external principal than without her.

Denote $C = \min\left(\frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}, k_H - 1\right)$. The following proposition describes the conditions under which the discussed cases are possible.

Proposition 3.

i) *For $\lambda < 1 + k_H$:*

if $\gamma \geq k_H - 1$ then the external principal always pays non-zero bonus in the interval $p \in [p^*, 1]$, the agent chooses the effort of 1, and the expected effort in the second period is $\bar{e} = 1 - p$;

if $\gamma < k_H - 1$ then there exists such p^γ that for all $p \in [p^*, p^\gamma]$ the external principal pays non-zero bonus, the agent chooses the effort of 1, and the expected effort in the second period is $\bar{e} = 1 - p$. If $p \in (p^\gamma, 1]$ then both principals pay zero bonuses.

ii) *For $\lambda \geq 1 + k_H$:*

if $\gamma > 0$ there exists such p^γ that for all $p \in [p^*, p^\gamma]$ the external principal pays a non-zero bonus, the agent chooses $e_1 = 1$, and the expected effort in the second period is $\bar{e}_2 = 1 - p$;

If $\gamma \leq C$ then for all $p \in [1/k_H, 1]$ both external and self-principal pay zero bonus in period 1;

If $\gamma > C$ then there exist an interval in $p \in [1/k_H, 1]$ such that the external principal pays a non-zero bonus, the agent chooses $e_1 = 1$, and the expected effort in the second period is $\bar{e}_2 = 1 - p$.

I prove the proposition in the Appendix. The intuition behind this proposition is the following. The decision of the external principal to pay a bonus depends on two factors: whether the self-principal decides to pay a bonus, and how much influence the external bonus has on the agent's incentive to work (parameter γ). Three cases are possible. If the agent's self-confidence is high ($p < p^*$), the self-principal will always pay a bonus in the first period, independently of whether the external principal is present. Therefore, external principal free rides and pays zero for any γ. In terms of efforts, the presence of the external principal does not have any effect on the self-control motivation, since the person is strong enough to abstain herself. The chances to abstain in the second period do not change as well.

If the self-confidence is intermediate or low, in equilibrium either both principals pay a zero bonus or a non-zero one. Whenever γ is sufficiently large (e.g., $\gamma > C$ or $\gamma \geq k_H - 1$ in the proposition), the external reward has a lot of influence on the agent's decision to abstain. Hence, the agent may choose $e_1 = 1$ even for a small external bonus. This makes payment profitable for both external and self-principal, since it reduces the bonuses that each of them has to pay. The presence of an external bonus makes the agent abstain, although he would not have abstained given the self-principal alone. In terms of efforts chosen, when the self-confidence is low (i.e. $p \in [1/k_H, 1]$ or $p \in [p^*, 1]$ in the proposition) the agent is more likely to exert positive effort in the long-run after the external principal than without her. On the contrary, when the self-confidence is not too low (i.e. $p \in [p^*, p^\gamma]$), then the agent will choose $e_1 = 1$ under external principal and $e_1 = 0$ without her. As a consequence, in the second period the expected effort will be $\bar{e}_2 = 1 - p$ after the external principal and $\bar{e}_2 = 1$ after the self-principal alone. This means, that in the short-run abstention is more likely with the external reward, while in the long-run the self-control is more likely without the preceding intervention.

In case γ is low, the external principal finds it too costly to stimulate abstention and pays zero. The self-principal also cannot afford a sufficiently high bonus and pays zero as well. Here, the presence of the external principal does not affect the agent's choice of effort.

In this setting I have considered the external principal that is strategic. She chooses the reward given the parameters of a particular individual and taking into account her willingness to abstain without an external reward. In reality this is not always the case. When an individual decides to participate in a self-control promoting program, the person or the organization offering a reward may not know the individual level of self-confidence or her sensitivity to an external bonus. Therefore, the bonus cannot be tailor-made for every participant of the program, but is fixed. It is easy to see that the case where the external principal does not choose a bonus strategically is a subcase of the model outlined above. The solution of the self-principal is the same, since in both cases she takes the external bonus as given. The cases where the strategic external principal would choose zero bonus are analogous to the situation in which the non-strategic principal offers an *insufficient* bonus to make a person abstain. High bonus paid by the non-strategic principal is identical to the case of the agent being highly sensitive to the reward of the strategic principal (large γ). The only difference is that the non-strategic principal may offer a very high bonus, such that the self-principal will choose to pay zero. Then, for any level of self-confidence the agent will exert effort in the first period, but in the second period only that person will maintain motivation for self-control who observed low abstention costs after the first period effort.

Next, I consider the extension to the benchmark case of the model. I analyze the model solution assuming that the external principal cares about the outcomes of both periods but may pay a bonus only in the first one. I formulate the following proposition[3]:

Denote $C_1 = \min\left(\frac{k_H}{2k_H-1}\left(\frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}\right), k_H - 1\right)$.

Proposition 4.

i) *For* $\lambda < 1 + k_H$*:*

if $\gamma \geq k_H - 1$ *then the external principal always pays non-zero bonus in the interval* $p \in [p^*, 1]$*, the agent chooses the effort of 1, and the expected effort in the second period is* $\bar{e} = 1 - p$*;*

[3] Note that I provide sufficient conditions and do not describe the solution in full as in Proposition 3. The reason for this is that the intuition is very similar to that of Proposition 3 yet the computation is more complicated. Hence, I concentrate on the most interesting cases that correspond to the results in Proposition 3.

if $\gamma < k_H - 1$ then there exists such p^γ that for all $p \in [p^*, p^\gamma]$ the external principal pays non-zero bonus, the agent chooses the effort of 1, and the expected effort in the second period is $\bar{e} = 1 - p$. If $p \in (p^\gamma, 1]$ then both principals pay zero bonuses.

ii) *For $\lambda \geq 1 + k_H$:*

if $\gamma > 0$ there exists such p^γ that for all $p \in [p^*, p^\gamma]$ the external principal pays a non-zero bonus, the agent chooses $e_1 = 1$, and the expected effort in the second period is $\bar{e}_2 = 1 - p$;

If $\gamma \leq C_1$ then for all $p \in [1/k_H, 1]$ both external and self-principal pay zero bonus in period 1;

If $\gamma > C_1$ then there exist an interval in $p \in [1/k_H, 1]$ such that the external principal pays a non-zero bonus, the agent chooses $e_1 = 1$, and the expected effort in the second period is $\bar{e}_2 = 1 - p$.

Note that the external principal's behavior described in this proposition is identical to that in Proposition 3: the only difference is in the definition of the constant C_1. Intuitively, if the external principal participates in both periods and cares about both outcomes, her choice is very similar to the one-period case if $\lambda < 1 + k_H$: the only difference is that the threshold p^γ decreases as the external principal gets utility from the second period.

However, if $\lambda \geq 1 + k_H$ then the trade-off for the external principal changes. For $p \in [p^*, 1/k_H]$ she gets 1 in the first period and $1 - p$ in the second when the agent exerts $e_1 = 1$. If $e_1 = 0$ then the agent will be motivated by the self-principal in the second period and $e_2 = 1$. The external principal gets $1 - p$ of additional utility if she motivates the agent to work in the first period. If $p \in (1/k_H, 1]$ then the agent will not exercise any effort in both periods on his own. Thus the additional utility the external principal gets from making the agent exert effort in the first period would be $2 - p$.

Note that the results for the long-lived and short-lived external principal are qualitatively the same. This is intuitive as the only difference in the external principal's utility function consists in the additional utility she receives from abstention in period 2. Yet the influence of her bonus b_{EP} on the agent is determined by γ. Varying this parameter we achieve qualitatively the same result in both cases. Therefore, the differences in the utility function do not qualitatively impact the external principal's behavior.

Conclusion

The theoretical model presented in this chapter allows to explain the contradictory empirical evidence regarding the positive and negative effects of rewards. This model demonstrates the intuition that is different from that described in other papers attempting to explain the phenomena of motivational crowding-in and crowding-out. Instead of being based on the leakage of private information from the principal to the agent (Bénabou and Tirole (2003)) or on the individual preferences for being internally or externally motivated (Harvey(2005)), the model uses the concept of a trade-off between achieving better short-run results by induction of higher efforts and being averse to revelation of information about abstention costs.

The developed model makes several contributions to the analysis of the self-control problem and motivational crowding out.

First, the model shows that motivational crowding out can occur even when both the principal and the agent have identical information on the agent's level of ability. The main driving force of the model is based on the fact that the larger is the effort level exerted by the agent the more information an individual gets about her abstention costs. If this information is negative, this may crowd out motivation to abstain in the future.

Second, the model allows to see that the behavior of people with high self-confidence regarding their self-control abilities is not particularly influenced by the presence of an external reward. On the contrary, for people with low self-confidence external reward may make abstention more likely than self-rewards. For those with the intermediate self-confidence external and self-bonuses demonstrate the opposing effects on the likelihood of abstention in the short- and in the long-term perspective.

Finally, the model may shed light on the fact that the property of motivational crowding-out is mostly a feature of rewards for achievement as opposed to rewards for participation. In the present paper I have mostly focused on the analysis of rewards delivered for a success in a certain activity requiring self-control. Deci and Ryan(1985), Charness and Gneezy (2008) and Jochelson (2007) suggests that monetary rewards offered for a mere participation in an activity often increase participation rates and do not have an undermining effect on motivation in a post-reward period. On the other hand, performance-contingent rewards are most likely to lead to the motivational crowding-out after the rewards are withdrawn. One possible explanation for this would be habit formation during the reward period. However, the presented model suggests a different view on this issue. The main problem with reward for achievement is that it establishes a direct link between efforts exerted and the payment received. Therefore, if an agent wants to get high payment he should exert high efforts which inevitably leads to revealing the cost of abstention and potential loss of motivation. If reward is paid merely for participation in an activity, then this link between payment and effort is no more present. Hence, an individual may still choose lower effort level and avoid revealing the costs at the same time preserving her self-confidence for the future period. This conclusion represents a testable prediction of the model. If it is possible to obtain an unambiguous measure of individual disappointment-aversion, then an experiment may demonstrate that people with higher degree of disappointment-aversion are more prone to decrease efforts under reward-for-participation scheme given that performing an activity may reveal certain useful information about the subject's skills. In this case motivational crowding-out should occur to a lesser extent in future periods compared to being rewarded for achievement. In the latter case reducing the efforts is more costly.

Appendix: Proof of Proposition 1

First, I calculate the utility function of the self-principal (8). The bonus paid to the agent in the first period is equal to $\hat{k}$. In the second period, agent is paid 0 regardless of the true cost. Expected outcome is equal to $p \times 0 + (1-p) \times 1 = (1-p)$ and hence from (5):

$$EU_1^{SP}(e_1 = 1) = (1-\hat{k}) + (1-p)(1+1-(1-p)) + \lambda p(0-(1-p)) =$$

$$= (1-\hat{k}) + (-\lambda p(1-p) + (1-p)(1+p)).$$

Assume first that $p < \bar{p}$. Given $e_1 = 0$, the self-principal pays bonus $\hat{k}$ in the second period.

In the beginning of period 1, the self-principal compares utilities $EU_1^{SP}(e_1 = 1)$ and $EU_1^{SP}(e_1 = 0) = 1 - \hat{k}$:

$$EU_1^{SP}(e_1 = 1) - EU_1^{SP}(e_1 = 0) = (-\lambda p(1-p) + (1-p)(1+p)). \tag{9}$$

Thus, the utility of the self-principal is larger with $e_1 = 1$ if $1 + p > \lambda p$, $p < \frac{1}{\lambda - 1}$.

There are two possible cases:

1) $\frac{1}{\lambda-1} < \bar{p} = \frac{1}{k_H}$ or $\lambda > 1 + k_H$. In this case, in the first period the self-principal pays bonus $\hat{k}$ if $p < p^* = \frac{1}{\lambda-1}$, and pays 0, otherwise.

2) $\frac{1}{\lambda-1} \geq \bar{p} = \frac{1}{k_H}$ or $\lambda \leq 1 + k_H$. In this case, the threshold $\frac{1}{\lambda-1}$ is larger than the point $\bar{p}$ after which the self-principal does not pay the agent in the second period. This means that for all $p < \bar{p}$, $EU_1^{SP}(e_1 = 1) - EU_1^{SP}(e_1 = 0) > 0$. Therefore, to derive p^* we should compare $EU_1^{SP}(e_1 = 1)$ to the utility function given $e_1 = 0$ in the interval $p \in \left[\bar{p}, \frac{1}{\lambda-1}\right]$ which is equal to $EU_1^{SP}(e_1 = 0, p \geq \bar{p}) = 0$.

Hence, I compare:

$$EU_1^{SP}(e_1 = 1) - EU_1^{SP}(e_1 = 0, p \geq \bar{p}) = p^2(\lambda - 1) - p(k_H + \lambda) + 2. \tag{10}$$

Note that the minimum of this function is achieved at $p = \frac{k_H+\lambda}{2(\lambda-1)} \geq 1$ as $2 + k_H > 1 + k_H \geq \lambda$. Moreover, the function is positive at $p = \bar{p}$ (note that it is proven that the function is larger than $1-\hat{k}$ in the interval $[0, \bar{p}]$) and it is negative at $p = 1$ as $1 - k_H < 0$. Thus there is a unique point at which the function crosses 0 in the interval $[\bar{p}, 1]$.

Solving $p^2(\lambda-1) - p(k_H+\lambda) + 2 = 0$ we get $p^* = \frac{k_H+\lambda-\sqrt{(k_H+\lambda)^2-8(\lambda-1)}}{4(\lambda-1)}$. Note that given $\lambda \leq 1 + k_H$:

$$(k_H+\lambda)^2-8(\lambda-1) \geq (2\lambda-1)^2-8(\lambda-1) = 4\lambda^2-4\lambda+1-8\lambda+8 = 4\lambda^2-12\lambda+9 = (2\lambda-3)^2 \geq 0.$$

Thus, the root is correctly defined. The second root, $\frac{k_H+\lambda+\sqrt{(k_H+\lambda)^2-8(\lambda-1)}}{4(\lambda-1)}$, is larger than the first one and is larger than 1.

If $p \geq p^*$ and $\lambda \leq 1+k_H$, the self-principal pays zero bonus in the first period. Q.E.D.

Appendix: Proof of Proposition 2

i) From the proof of Proposition 1 we know that in the first period, the self-principal pays bonus $\hat{k}$ if $p \leq p^*$ and pays zero, otherwise. This means that $e_1 = 1$ if $p \leq p^*$, and in the second period only the agent with low cost exerts effort. Thus, the expected effort is $\bar{e}_2 = 1 - p$.

If $p > p^*$ then $e_1 = 0$ and there is no revealing. Moreover, as $p^* > \bar{p}$, the self-principal pays zero bonus in the second period. Hence $\bar{e}_2 = 0$.

ii) From the proof of Proposition 1, in the first period the self-principal pays bonus $\hat{k}$ if $p \leq p^*$. Thus $e_1 = 1$, there is revealing and only the agents with low costs work in the second period which means that $\bar{e}_2 = 1 - p$.

If $p \in (p^*, 1/k_H]$ then the self-principal pays zero in the first period but pays $\hat{k}$ in the second period. This means that no agent works in the first period and all of them work in the second period, so $e_1 = 0$ and $\bar{e}_2 = 1$.

If $p > 1/k_H$ then the self-principal pays zero in both periods and thus $e_1 = 0$, $\bar{e}_2 = 0$.

Appendix: Proof of Proposition 3

To prove the Proposition 3 I first should calculate the optimal behavior of the external principal in different intervals of parameters.

As is shown in Proposition 1, there are two major cases: $\lambda > 1+k_H$ and $1 < \lambda \leq 1+k_H$. I start the solution with the second case.

Case 1: $1 < \lambda \leq 1 + k_H$.

We know from the proof of Proposition 1 that self-principal pays the agent to apply effort $e_1 = 1$ if $p \leq p^*$ where $p^* > 1/k_H$. This means that the external principal pays zero bonus in this interval.

If $p \in (p^*, 1]$ then the self-principal does not motivate agent to work in the first period because the self-principal gets negative utility. Consider the utility function of the self-principal in this interval given bonus b_{EP} of the external principal:

$$EU_1^{SP}(e_1 = 1) = \gamma b_{EP} + (1 - \hat{k}) + (-\lambda p(1-p) + (1-p)(1+p)).$$

This relation is derived from the utility function of the agent: the external principal's bonus reduces the self-principal's bonus by γb_{EP}. External principal should pay a bonus such that the self-principal break-even. This means that

$$\gamma b_{EP} = -1 + \hat{k} + \lambda p(1-p) - (1-p)(1+p). \tag{11}$$

Moreover, this bonus should be no larger than 1 and thus we have

$$\gamma \geq -1 + \hat{k} + \lambda p(1-p) - (1-p)(1+p). \tag{12}$$

Denote $f(p) = -1 + \hat{k} + \lambda p(1-p) - (1-p)(1+p) = -p^2(\lambda - 1) + p(\lambda + k_H) - 2$. This is a quadratic function in p and it crosses zero at $p = p^*$. We can compute the point of the maximum value of this function:

$$f'(p) = -2(\lambda - 1)p + (\lambda + k_H) = 0, \, p_1 = \frac{\lambda + k_H}{2(\lambda - 1)}.$$

Here p_1 is the point at which the maximum for $f(p)$ is achieved. Note that we have $\lambda \leq 1 + k_H$ and thus

$$p_1 = \frac{\lambda + k_H}{2(\lambda - 1)} \geq \frac{\lambda + \lambda - 1}{2\lambda - 2} > 1.$$

This means that the unrestricted maximum is achieved outside the interval $[p^*, 1]$ and the maximum value in this interval is achieved at $p = 1$ because $f(p^*) = 0$ and $p_1 > 1$. I compute $f(1) = -(\lambda - 1) + \lambda + k_H - 2 = k_H - 1$. This function is increasing in the interval $[p^*, 1]$.

In the case $\gamma \geq k_H - 1$ the only possible point of intersection between the function $f(p)$ and the function $g(p) = \gamma$ in the interval $[p^*, 1]$ is $p = 1$. This means that the external principal may always pay a bonus enough to make agent apply effort $e_1 = 1$.

In the case $\gamma < k_H - 1$ the functions $f(p)$ and $g(p)$ intersect at the point solving

$$-p^2(\lambda - 1) + p(\lambda + k_H) - 2 = \gamma, \, p^{**} = \frac{\lambda + k_H - \sqrt{(\lambda + k_H)^2 - 4(2 + \gamma)(\lambda - 1)}}{2(\lambda - 1)}$$

(the other root is greater than 1 as $\lambda + k_H > 2(\lambda - 1)$). The external principal compensates for the self-principal's bonus until the point p^{**} and both principals pay zero if $p \in [p^{**}, 1]$.

Thus in the interval $p \in [p^*, 1]$ the solution is the following:

i. If $\gamma \geq k_H - 1$ then the external principal pays the bonus

$$b_{EP} = \frac{-p^2(\lambda - 1) + p(\lambda + k_H) - 2}{\gamma}, \tag{13}$$

the agent supplies effort $e_1 = 1$ in the first period, and the self-principal pays $\hat{k} - b_{EP}$.

ii. If $\gamma < k_H - 1$ and $p \in [p^*, p^{**}]$ then the external principal pays the bonus (22), the agent supplies effort $e_1 = 1$ in the first period, and the self-principal pays $\hat{k} - b_{EP}$.

If $\gamma < k_H - 1$ and $p \in (p^{**}, 1]$ then both principals pay zero bonus and the agent's effort is zero in the first period.

This concludes case 1.

Case 2: $\lambda > 1 + k_H$.

In this case, as I proved in Proposition 1, $p^* < 1/k_H$ and the self-principal pays the bonus if $p \leq p^*$. Hence the external principal pays zero bonus in this interval. If $p > p^*$, the self-principal pays zero bonus.

Consider first the interval $p \in [p^*, 1/k_H]$. In this interval the difference between utilities the self-principal gets if the effort in the first period is 1 or 0 is given by (9). Denote

$$f_1(p) = -(-\lambda p(1-p) + (1-p)(1+p)) = -p^2(\lambda - 1) + \lambda p - 1.$$

This function achieves its maximum at $p_{21} = \frac{\lambda}{2(\lambda-1)}$ and the (unrestricted) maximum is equal to $f_1^{\max} = \frac{\lambda^2}{2(\lambda-1)} - \frac{\lambda^2}{4(\lambda-1)} - 1 = \frac{(\lambda-2)^2}{4(\lambda-1)}$.

Consider now the interval $p \in [1/k_H, 1]$. In this interval the difference between utilities the self-principal gets if the effort in the first period is 1 or 0 is given by (10). Denote

$$f_2(p) = -(p^2(\lambda - 1) - p(k_H + \lambda) + 2) = -p^2(\lambda - 1) + p(\lambda + k_H) - 2.$$

Note that $f_2(p) \equiv f(p)$ and thus has the same point of maximum: $p_{22} = \frac{\lambda + k_H}{2(\lambda-1)}$. However, the parameters are different in the Case 2. The value at this point is

$$f_2^{\max} = \frac{(\lambda + k_H)^2}{2(\lambda - 1)} - \frac{(\lambda + k_H)^2}{4(\lambda - 1)} - 2 = \frac{(\lambda + k_H)^2}{4(\lambda - 1)} - 2.$$

The external principal should add a bonus to the self-principal payment in order to make agent work in the first period. This means that in the interval $p \in [p^*, 1/k_H]$ the agent exerts effort $e_1 = 1$ only if

$$\gamma b_{EP} = f_1(p), \gamma \geq f_1(p). \tag{14}$$

Similarly, in the interval $p \in [1/k_H, 1]$ the agent exerts effort $e_1 = 1$ only if

$$\gamma b_{EP} = f_2(p), \gamma \geq f_2(p). \tag{15}$$

The two functions, $f_1(p)$ and $f_2(p)$, are both quadratic. Their points of maximum may either lay in the respective intervals or to the left/right of these intervals. I next study their positions depending on the parameters λ and k_H.

Let's start with p_{21}. This point should be compared with $p^* = \frac{1}{\lambda-1}$ and $\frac{1}{k_H}$. We have

$$p_{21} \leq p^* \Leftrightarrow \frac{\lambda}{2(\lambda-1)} \leq \frac{1}{\lambda-1} \Leftrightarrow \lambda \leq 2.$$

Yet in the Case 2, $\lambda > 1 + k_H > 2$. Hence $p_{21} > p^*$ always.

Next, we compare

$$p_{21} \leq \frac{1}{k_H} \Leftrightarrow \frac{\lambda}{2(\lambda-1)} \leq \frac{1}{k_H} \Leftrightarrow k_H\lambda \leq 2(\lambda-1).$$

If $k_H \geq 2$, this cannot hold and thus $p_{21} > \frac{1}{k_H}$ always. If $k_H < 2$ then

$$p_{21} \leq \frac{1}{k_H} \Leftrightarrow \lambda \geq \frac{2}{2-k_H} \equiv \lambda_1. \tag{16}$$

Now we consider p_{22}. We have

$$p_{22} \leq \frac{1}{k_H} \Leftrightarrow \frac{\lambda + k_H}{2(\lambda-1)} \leq \frac{1}{k_H} \Leftrightarrow k_H\lambda + k_H^2 \leq 2(\lambda-1).$$

If $k_H \geq 2$, this cannot hold and thus $p_{22} > \frac{1}{k_H}$ always. If $k_H < 2$ then

$$p_{22} \leq \frac{1}{k_H} \Leftrightarrow \lambda \geq \frac{2+k_H^2}{2-k_H} \equiv \lambda_2. \tag{17}$$

Note that $\lambda_2 > \lambda_1$ if $k_H < 2$ as the numerator of the former is larger.

Compare p_{22} to 1:

$$p_{22} \leq 1 \Leftrightarrow \frac{\lambda + k_H}{2(\lambda-1)} \leq 1 \Leftrightarrow \lambda \geq k_H + 2 \equiv \lambda_3.$$

Note that if $k_H < 2$ then

$$\frac{2+k_H^2}{2-k_H} - (2+k_H) = \frac{2+k_H^2 - 4 + k_H^2}{2-k_H} > 0.$$

Thus $\lambda_2 > \lambda_3$ if $k_H < 2$.

Note also that $\lambda_1 > 1 + k_H$ and $\lambda_3 > 1 + k_H$ if $k_H < 2$:

$$\lambda_1 > 1 + k_H \Leftrightarrow \frac{2}{2-k_H} > 1 + k_H \Leftrightarrow \frac{2-2+k_H-2k_H+k_H^2}{2-k_H} = \frac{k_H^2 - k_H}{2-k_H} > 0,$$

$$\lambda_3 > 1 + k_H \Leftrightarrow k_H + 2 > k_H + 1.$$

Compare λ_1 and λ_3 assuming $k_H < 2$:

$$\lambda_1 \leq \lambda_3 \Leftrightarrow \frac{2}{2-k_H} - (2+k_H) = \frac{k_H^2 - 2}{2-k_H} \leq 0.$$

Hence $\lambda_1 \leq \lambda_3$ if $k_H \in (1, \sqrt{2}]$, and $\lambda_1 > \lambda_3$ if $k_H \in (\sqrt{2}, 2)$. We can now summarize the cases for p_{21} and p_{22}.

A. If $k_H \in (1, \sqrt{2}]$ then $\lambda_1 \leq \lambda_3 < \lambda_2$, and:

A.1. If $\lambda \in (1 + k_H, \lambda_1]$ then $p_{21} > 1/k_H$ and $p_{22} \geq 1$.

A.2. If $\lambda \in (\lambda_1, \lambda_3]$ then $p_{21} \leq 1/k_H$ and $p_{22} \geq 1$.

A.3. If $\lambda \in (\lambda_3, \lambda_2]$ then $p_{21} \le 1/k_H$ and $1/k_H \le p_{22} \le 1$.

A.4. If $\lambda > \lambda_2$ then $p_{21} \le 1/k_H$ and $p_{22} < 1/k_H$.

B. If $k_H \in (\sqrt{2}, 2)$ then $\lambda_3 < \lambda_1 < \lambda_2$, and:

B.1. If $\lambda \in (1 + k_H, \lambda_3]$ then $p_{21} > 1/k_H$ and $p_{22} \ge 1$.

B.2. If $\lambda \in (\lambda_3, \lambda_1]$ then $p_{21} > 1/k_H$ and $1/k_H \le p_{22} \le 1$.

B.3. If $\lambda \in (\lambda_1, \lambda_2]$ then $p_{21} \le 1/k_H$ and $1/k_H \le p_{22} \le 1$.

B.4. If $\lambda > \lambda_2$ then $p_{21} \le 1/k_H$ and $p_{22} < 1/k_H$.

C. If $k_H \ge 2$ then $p_{21} > 1/k_H$, and:

C.1. If $\lambda \in (1 + k_H, \lambda_3]$ then $p_{22} \ge 1$.

C.2. If $\lambda > \lambda_3$ then $1/k_H \le p_{22} \le 1$.

The summary above shows what happens in each case and in which point functions f_1, f_2 achieve their maximums. For example, in the case A.1 the maximum value of f_1 in the interval $[p^*, 1/k_H]$ is at $p = 1/k_H$ and the maximum value of f_2 in the interval $[1/k_H, 1]$ is at $p = 1$. In the case A.3, however, the maximum values are achieved at p_{21} and p_{22}, respectively.

We need also to calculate $f_1(1/k_H)$, $f_2(1/k_H)$ and $f_2(1)$ to find the maximum values in the cases where p_{21} and p_{22} lie outside the intervals $[p^*, 1/k_H]$ and $[1/k_H, 1]$, respectively. I compute

$$f_1(1/k_H) = f_2(1/k_H) = \frac{\lambda}{k_H} - 1 - \frac{\lambda - 1}{k_H^2}, \; f_2(1) = k_H - 1.$$

I calculate below the maximum values in every case A.1-C.2 for f_1 (denote it g_1) and f_2 (denote it g_2:

A.1 $g_1 = \frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}$, $g_2 = k_H - 1$.

A.2 $g_1 = f_1^{\max}$, $g_2 = k_H - 1$.

A.3 $g_1 = f_1^{\max}$, $g_2 = f_2^{\max}$.

A.4 $g_1 = f_1^{\max}$, $g_2 = \frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}$.

B.1 $g_1 = \frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}$, $g_2 = k_H - 1$.

B.2 $g_1 = \frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}$, $g_2 = f_2^{\max}$.

B.3 $g_1 = f_1^{\max}$, $g_2 = f_2^{\max}$.

B.4 $g_1 = f_1^{\max}$, $g_2 = \frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}$.

C.1 $g_1 = \frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}$, $g_2 = k_H - 1$.

C.2 $g_1 = \frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}$, $g_2 = f_2^{\max}$.

In the cases A.1, B.1, B.2, C.1 and C.2 we have $g_1 \le g_2$. This is derived from the fact that when $p_{22} \ge 1$ then $f_2(1) > f_2(1/k_H) = f_1(1/k_H)$ (cases A.1, B.1 and C.1), while in the other two cases (B.2 and C.2) the maximum of f_2 over the interval $[1/k_H, 1]$ is no less than $f_2(1/k_H)$.

In the cases A.4 and B.4 we observe that the maximum of f_1 in the interval $[p^*, 1/k_H]$ is no less than $f_1(1/k_H)$ which coincides with g_2. Thus in these cases $g_1 \ge g_2$.

For the rest of the cases with two internal maximums it is not straightforward to determine the conditions relating g_1 and g_2. In these cases I will explain the behavior of agents and principals without giving explicit condition relating g_1, g_2.

I derive now the points in which f_1 and f_2 cross constant γ. Note that if $\gamma \ge g_1$ ($\gamma \ge g_2$) then f_1 (f_2) lies below γ in the respective interval. Consider the case $\gamma < g_1$ (note that this condition depends on g_1 and is different for each scenario A.1-C.2). We may have no more than two points of intersection between γ and f_1 in the interval $[p^*, 1/k_H]$:

$$\gamma = -p^2(\lambda-1) + \lambda p - 1,$$

$$p_{L1}^{\gamma} = \frac{\lambda - \sqrt{\lambda^2 - 4(1+\gamma)(\lambda-1)}}{2(\lambda-1)},$$

$$p_{L2}^{\gamma} = \frac{\lambda + \sqrt{\lambda^2 - 4(1+\gamma)(\lambda-1)}}{2(\lambda-1)}.$$

Here upper index γ means that I solve for the crossing between f_1 and γ, lower index L means that the point is in the "left" interval $[p^*, 1/k_H]$, and lower indices 1,2 mean that we consider the root with minus or plus (the second one is obviously larger).

To choose which root lies in the interval $[p^*, 1/k_H]$ we follow a simple rule. Function f_1 starts at zero ($f_1(p^*) = 0$) and increases until its "peak" (point at which the maximum is reached), then it decreases. If the peak lies outside the interval (cases A.1, B.1, C.1 and C.2) then there is only one point at which γ and f_1 intersect in the interval, and this point is p^{γ}_{L1} as the second point is larger. If the peak is inside the interval then there are two points of intersection, both p^{γ}_{L1} and p^{γ}_{L2}, but only if $\gamma > f_1(1/k_H)$ - otherwise there is only one point of intersection p^{γ}_{L1}.

For the "right" interval $p \in [1/k_H, 1]$ we solve $\gamma = f_2(p)$,

$$\gamma = -p^2(\lambda - 1) + (\lambda + k_H)p - 2,$$

$$p^{\gamma}_{R1} = \frac{\lambda + k_H - \sqrt{(\lambda + k_H)^2 - 4(2+\gamma)(\lambda - 1)}}{2(\lambda - 1)},$$

$$p^{\gamma}_{R2} = \frac{\lambda + k_H + \sqrt{(\lambda + k_H)^2 - 4(2+\gamma)(\lambda - 1)}}{2(\lambda - 1)}.$$

In this interval we have a different picture as the peak of f_2 may lie to the left, to the right and in the interval, and also minimum value of f_2 is achieved at either $1/k_H$ or 1. If γ is less than this minimum then the external principal pays zero bonus. Assuming $f_2^{\min} \le \gamma < g_2$ we get the following:

In the cases A.1, A.2, B.1 and C.1 the peak is larger than 1, f_2 is increasing in the interval and there is only one point of intersection p^{γ}_{R1}; the minimum value of f_2 is $f_2(1/k_H) = f_1(1/k_H)$.

In the cases A.3, B.2, B.3 and C.2 the peak is inside the interval and there are two points of intersection, p^{γ}_{R1} and p^{γ}_{R2}; the minimum value of f_2 may be either of two $f_2(1/k_H), f_2(1)$.

In the cases A.4 and B.4 the peak is lower than $1/k_H$, f_2 is decreasing in the interval and there is only one point of intersection p^{γ}_{R2}; the minimum value of f_2 is $f_2(1)$.

There are seven cases mentioned above in which we can directly compare g_1 and g_2 (A.1, A.4, B.1, B.2, B.4, C.1 and C.4). This is enough to explicitly state the relation between γ and the intervals in which the external principal pays bonus in the first period, and agent's effort is 1. In other three cases (A.2, A.3, B.3) I describe necessary and sufficient conditions to get the relation.

Note that the bonus paid (if non-zero) in the interval $p \in [p^*, 1/k_H]$ is equal to:

$$b_{EP}^L = \frac{-p^2(\lambda - 1) + p\lambda - 1}{\gamma}, \tag{18}$$

and the bonus paid in the interval $p \in [1/k_H, 1]$ is equal to:

$$b_{EP}^R = \frac{-p^2(\lambda - 1) + p(\lambda + k_H) - 2}{\gamma}, \tag{19}$$

Note that cases A.1, B.1 and C.1 correspond to the Figure 1, cases B.2 and C.2 - to the Figure 2, cases A.3 and B.3 - to the Figure 5, cases A.4 and B.4 - to the Figure 3, and case A.2 - to the Figure 4.

In the Table 1 I describe what happens in each of the cases A.1-C.2. I do not consider cases A.3 and B.3 because there are 20 possible combinations of parameters that define the intervals for bonuses in these two cases, and they do not add anything different to the intuition. Each row corresponds to up to three cases. Each column is devoted to the conditions on γ and other parameters that lead to the result given. Every cell contains the information on the intervals in which bonuses b_{EP}^L, b_{EP}^R or zero are paid. "n/a" means that this bonus is not paid in any interval.

The results in the Table 1 directly follow from the Figures 1-5. Namely, as γ decreases we have more intervals in which the external principal does not participate (pays zero bonus). These intervals are found from the intersection of γ with respective functions f_1, f_2 (see p_{L1}^γ, p_{L2}^γ, p_{R1}^γ, p_{R2}^γ).

Cases A.1, B.1 and C.1, as well as A.4 and B.4, are easier to describe because there is no uncertainty in the Figures 1 and 3: the respective positions of the values g_1, $f_1(1/k_H)$ and $f_1(1)$ are unique. Thus there are only 3 intervals for γ in the first three cases, and there are only 4 intervals for γ in the last two cases.

Where the external principal pays a bonus different from zero, the agent would exert effort of 1. This means that the Table 1 helps us to find the intervals in which crowding-out or crowding-in appears.

I can now describe the intervals of crowding-out and crowding-in.

I consider two situations of interest: 3.1) the agent works less in the second period after the external principal than he would do after the self-principal only; 3.2) the agent works more after the external principal than he would do after the self-principal only.

Situation 3.1 appears, for example, in the cases A.1, B.1 and C.1. In the interval $[p^*, 1/k_H]$ for large enough γ, or in the interval $[p^*, p^{\gamma}_{L1}]$ when γ is small enough, the external principal participates with the bonus b^L_{EP} and makes agent work in the first period. In the second period only the agents with low costs will still work. However, in this same interval, given $\gamma = 0$ the self-principal does not force agent to work in the first period yet pays him the bonus $\hat{k}$ in the second period. This means that the agent decreases his effort with respect to the case with no external principal. Note that this happens for any $\gamma > 0$ however small it is.

Situation 3.2 appears in the "right" interval $p \geq 1/k_H$. In this interval the self-principal never pays the agent enough to make him work, but the external principal may add a bonus to force the agent to work (in almost all the cases). Only low γ may prevent the external principal to participate: if $\gamma \leq \min(f_2(1/k_H), f_2(1))$ then the self-principal produces the same outcome as the combination of two principals (agent never works if $p > 1/k_H$).

Appendix: Proof of Proposition 4

To prove the Proposition 4 I first should calculate the optimal behavior of the external principal in the different intervals of parameters.

In this section I assume that the external principal cares about both periods. Thus, given Propositions 1-2, I can characterize the utility of the external principal.

As is shown in Proposition 1, there are two major cases: $\lambda > 1+k_H$ and $1 < \lambda \leq 1+k_H$. I start the solution with the second case.

Table 1: Intervals of external principal's payments

	$\gamma \geq g_2 \geq g_1$	$g_2 > \gamma \geq g_1$	$g_1 > \gamma \geq 0$	
A.1, B.1, C.1	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, 1]$ 0: $[0, p^*]$	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, p^\gamma_{R1}]$ 0: $[0, p^*]$ 0: $[p^\gamma_{R1}, 1]$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^R_{EP}: n/a 0: $[0, p^*]$ 0: $[p^\gamma_{L1}, 1]$	
	$\gamma \geq g_1 \geq g_2$	$g_1 > \gamma \geq g_2$	$g_2 > \gamma \geq f_2(1)$	$f_2(1) > \gamma > 0$
A.4, B.4	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, 1]$ 0: $[0, p^*]$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^L_{EP}: $[p^\gamma_{L2}, 1/k_H]$ b^R_{EP}: $[1/k_H, 1]$ 0: $[0, p^*]$ 0: $(p^\gamma_{L1}, p^\gamma_{L2})$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^R_{EP}: $[p^\gamma_{R2}, 1]$ 0: $[0, p^*]$ 0: $(p^\gamma_{L1}, p^\gamma_{R2})$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^R_{EP}: n/a 0: $[0, p^*]$ 0: $[p^\gamma_{L1}, 1]$
	$\gamma \geq g_2 \geq g_1$ $f_2(1/k_H) \geq f_2(1)$	$g_2 > \gamma \geq f_2(1/k_H)$ $f_2(1/k_H) \geq f_2(1)$	$f_2(1/k_H) > \gamma \geq f_2(1)$ $f_2(1/k_H) \geq f_2(1)$	$f_2(1) > \gamma > 0$ $f_2(1/k_H) \geq f_2(1)$
B.2, C.2	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, 1]$ 0: $[0, p^*]$	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, p^\gamma_{R1}]$ b^R_{EP}: $[p^\gamma_{R2}, 1]$ 0: $[0, p^*]$ 0: $(p^\gamma_{R1}, p^\gamma_{R2})$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^R_{EP}: $[p^\gamma_{R2}, 1]$ 0: $[0, p^*]$ 0: $[p^\gamma_{L1}, p^\gamma_{R2}]$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^R_{EP}: n/a 0: $[0, p^*]$ 0: $[p^\gamma_{L1}, 1]$
	$\gamma \geq g_2 \geq g_1$ $f_2(1/k_H) < f_2(1)$	$g_2 > \gamma \geq f_2(1)$ $f_2(1/k_H) < f_2(1)$	$f_2(1) > \gamma \geq f_2(1/k_H)$ $f_2(1/k_H) < f_2(1)$	$f_2(1/k_H) > \gamma > 0$ $f_2(1/k_H) < f_2(1)$
B.2, C.2	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, 1]$ 0: $[0, p^*]$	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, p^\gamma_{R1}]$ b^R_{EP}: $[p^\gamma_{R2}, 1]$ 0: $[0, p^*]$ 0: $(p^\gamma_{R1}, p^\gamma_{R2})$	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, p^\gamma_{R1}]$ 0: $[0, p^*]$ 0: $(p^\gamma_{R1}, 1]$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^R_{EP}: n/a 0: $[0, p^*]$ 0: $[p^\gamma_{L1}, 1]$
	$\gamma \geq g_1 \geq f_2(1)$	$g_1 > \gamma \geq f_2(1)$	$f_2(1) > \gamma \geq f_2(1/k_H)$	$f_2(1/k_H) > \gamma > 0$
A.2	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, 1]$ 0: $[0, p^*]$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^L_{EP}: $[p^\gamma_{L2}, 1/k_H]$ b^R_{EP}: $[1/k_H, 1]$ 0: $[0, p^*]$ 0: $(p^\gamma_{L1}, p^\gamma_{L2})$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^L_{EP}: $[p^\gamma_{L2}, 1/k_H]$ b^R_{EP}: $[1/k_H, p^\gamma_{R1}]$ 0: $[0, p^*]$ 0: $(p^\gamma_{L1}, p^\gamma_{L2})$ 0: $(p^\gamma_{R1}, 1]$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^R_{EP}: n/a 0: $[0, p^*]$ 0: $(p^\gamma_{L1}, 1]$
	$\gamma \geq f_2(1) > g_1$	$f_2(1) > \gamma \geq g_1$	$g_1 > \gamma \geq f_2(1/k_H)$	$f_2(1/k_H) > \gamma > 0$
A.2	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, 1]$ 0: $[0, p^*]$	b^L_{EP}: $[p^*, 1/k_H]$ b^R_{EP}: $[1/k_H, p^\gamma_{R1}]$ 0: $[0, p^*]$ 0: $(p^\gamma_{R1}, 1]$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^L_{EP}: $[p^\gamma_{L2}, 1/k_H]$ b^R_{EP}: $[1/k_H, p^\gamma_{R1}]$ 0: $[0, p^*]$ 0: $(p^\gamma_{L1}, p^\gamma_{L2})$ 0: $(p^\gamma_{R1}, 1]$	b^L_{EP}: $[p^*, p^\gamma_{L1}]$ b^R_{EP}: n/a 0: $[0, p^*]$ 0: $(p^\gamma_{L1}, 1]$

Case 1: $1 < \lambda \leq 1 + k_H$.

We know from the proof of Proposition 1 that self-principal pays the agent to apply effort $e_1 = 1$ if $p \leq p^*$ where $p^* > 1/k_H$. This means that the external principal pays zero bonus in this interval.

If $p \in (p^*, 1]$ then the self-principal does not motivate agent to work in the first period because the self-principal gets negative utility. Consider the utility function of the self-principal in this interval given bonus b_{EP} of the external principal:

$$EU_1^{SP}(e_1 = 1) = \gamma b_{EP} + (1 - \hat{k}) + (-\lambda p(1-p) + (1-p)(1+p)).$$

This relation is derived from the utility function of the agent: the external principal's bonus reduces the self-principal's bonus by γb_{EP}. External principal should pay a bonus such that the self-principal break-even. This means that

$$\gamma b_{EP} = -1 + \hat{k} + \lambda p(1-p) - (1-p)(1+p). \tag{20}$$

Note that the utility the external principal gets in the case $e_1 = 0$ is zero, while if the agent works in the first period then the abilities are revealed and fraction $1 - p$ of agents work in the second period. Thus $b_{EP} \leq 1 + 1 - p = 2 - p$, hence

$$\gamma(2-p) \geq -1 + \hat{k} + \lambda p(1-p) - (1-p)(1+p). \tag{21}$$

Denote $f(p) = -1 + \hat{k} + \lambda p(1-p) - (1-p)(1+p) = -p^2(\lambda - 1) + p(\lambda + k_H) - 2$. This is a quadratic function in p and it crosses zero at $p = p^*$. We can compute the point of the maximum value of this function:

$$f'(p) = -2(\lambda - 1)p + (\lambda + k_H) = 0, \, p_1 = \frac{\lambda + k_H}{2(\lambda - 1)}.$$

Here p_1 is the point at which the maximum for $f(p)$ is achieved. Note that we have

$\lambda \le 1 + k_H$ and thus

$$p_1 = \frac{\lambda + k_H}{2(\lambda - 1)} \ge \frac{\lambda + \lambda - 1}{2\lambda - 2} > 1.$$

This means that the unrestricted maximum is achieved outside the interval $[p^*, 1]$ and the maximum value in this interval is achieved at $p = 1$ because $f(p^*) = 0$ and $p_1 > 1$. I compute $f(1) = -(\lambda - 1) + \lambda + k_H - 2 = k_H - 1$. This function is increasing in the interval $[p^*, 1]$.

Function $g(p) = \gamma(2 - p)$ is decreasing in the interval $[p^*, 1]$. Its value at $p = p^*$ is positive ($g(p^*) = \gamma(2 - p^*)$) and its value at $p = 1$ is equal to $g(1) = \gamma$.

In the case $\gamma \ge k_H - 1$ the only possible point of intersection between the function $f(p)$ and the function $g(p)$ in the interval $[p^*, 1]$ is $p = 1$. This means that the external principal may always pay a bonus enough to make agent apply effort $e_1 = 1$.

In the case $\gamma < k_H - 1$ the functions $f(p)$ and $g(p)$ intersect at the point solving

$$-p^2(\lambda - 1) + p(\lambda + k_H) - 2 = \gamma(2 - p),\ p^{**} = \frac{\lambda + k_H + \gamma - \sqrt{(\lambda + k_H + \gamma)^2 - 4(2 + 2\gamma)(\lambda - 1)}}{2(\lambda - 1)}$$

(the other root is greater than 1 as $\lambda + k_H > 2(\lambda - 1)$). The external principal compensates for the self-principal's bonus until the point p^{**} and both principals pay zero if $p \in [p^{**}, 1]$.

Thus in the interval $p \in [p^*, 1]$ the solution is the following:

i. If $\gamma \ge k_H - 1$ then the external principal pays the bonus

$$b_{EP} = \frac{-p^2(\lambda - 1) + p(\lambda + k_H) - 2}{\gamma}, \tag{22}$$

the agent supplies effort $e_1 = 1$ in the first period, and the self-principal pays $\hat{k} - b_{EP}$.

ii. If $\gamma < k_H - 1$ and $p \in [p^*, p^{**}]$ then the external principal pays the bonus (22), the agent supplies effort $e_1 = 1$ in the first period, and the self-principal pays $\hat{k} - b_{EP}$.

If $\gamma < k_H - 1$ and $p \in (p^{**}, 1]$ then both principals pay zero bonus and the agent's effort is zero in the first period.

This concludes case 1.

Case 2: $\lambda > 1 + k_H$.

In this case, as I proved in Proposition 1, $p^* < 1/k_H$ and the self-principal pays the bonus if $p \leq p^*$. Hence the external principal pays zero bonus in this interval. If $p > p^*$, the self-principal pays zero bonus on his own.

Consider first the interval $p \in [p^*, 1/k_H]$. In this interval the difference between utilities the self-principal gets if the effort in the first period is 1 or 0 is given by (9). Denote

$$f_1(p) = -(-\lambda p(1-p) + (1-p)(1+p)) = -p^2(\lambda - 1) + \lambda p - 1.$$

The external principal gets $2 - p$ if $e_1 = 1$ as only the fraction $1 - p$ of agents works in the second period. If $e_1 = 0$, the self-principal motivates the agent to work in the second period and thus the external principal gets 1. This means that the maximum bonus $b_{EP} = 1 - p$.

The external principal should add a bonus to the self-principal payment in order to make agent work in the first period. This means that in the interval $p \in [p^*, 1/k_H]$ the agent exerts effort $e_1 = 1$ only if

$$\gamma b_{EP} = f_1(p), \gamma(1-p) \geq f_1(p). \tag{23}$$

This means the bonus paid is equal to

$$b_{EP} = \frac{-p^2(\lambda - 1) + \lambda p - 1}{\gamma}. \tag{24}$$

Consider now the interval $p \in [1/k_H, 1]$. In this interval the difference between utilities the self-principal gets if the effort in the first period is 1 or 0 is given by (10). Denote

$$f_2(p) = -(p^2(\lambda - 1) - p(k_H + \lambda) + 2) = -p^2(\lambda - 1) + p(\lambda + k_H) - 2.$$

The external principal gets $2-p$ if $e_1 = 1$ as only the fraction $1-p$ of agents works in the second period. If $e_1 = 0$, the self-principal cannot motivate the agent to work in the second period and thus the external principal gets 0. This means that the maximum bonus $b_{EP} = 2-p$.

Hence in the interval $p \in [1/k_H, 1]$ the agent exerts effort $e_1 = 1$ only if

$$\gamma b_{EP} = f_2(p), \gamma(2-p) \geq f_2(p). \tag{25}$$

The two functions, $f_1(p)$ and $f_2(p)$, are both quadratic. Their points of maximum are studied in the proof of Proposition 3.

In this proposition I compare f_1 to $h_1(p) = \gamma(1-p)$ and f_2 to $h_2(p) = \gamma(2-p)$. Note that the incentives of the external principal have changed: before the comparison was uniform (f_1 and f_2 to γ). Now, the external principal has more incentives to motivate agent in the right interval $[1/k_H, 1]$.

Both functions h_1 and h_2 are linear and decreasing. I start the analysis from f_1, h_1.

Note that $f_1(1) = h_1(1) = 0$. Thus the only possible tangency line to f_1 of the form $\gamma(1-p)$ is the one that has one point of intersection at $p = 1$. Any other line h_1 crosses the graph f_1 twice in the interval $[p^*, 1]$ - at the point $p = 1$ and at the point p^{γ}_{L1}:

$$-p^2(\lambda-1)+\lambda p-1 = \gamma(1-p), -p^2(\lambda-1)+(\lambda+\gamma)p-1-\gamma = 0, (p-1)(-p(\lambda-1)+1+\gamma) = 0;$$

$$p^{\gamma}_{L1} = \frac{1+\gamma}{\lambda-1}.$$

This point is larger than $p^* = \frac{1}{\lambda-1}$ and should be less than $1/k_H$ or

$$\frac{1+\gamma}{\lambda-1} \leq \frac{1}{k_H}, \gamma \leq \frac{\lambda-1}{k_H} - 1.$$

So with a small γ, f_1 and h_1 cross inside the interval $[p^*, 1/k_H]$. The functions cross after $1/k_H$ if γ is large enough.

This leads to the conclusion: if $0 < \gamma \le \frac{\lambda-1}{k_H} - 1$ then the external principal pays bonus 18 in the interval $[p^*, p^\gamma_{L1}]$, the agents works $e_1 = 1$ and $\mathbf{E}e_2 = 1 - p$; in the interval $(p^*, p^\gamma_{L1}]$ the external principal pays zero bonus, the agents works $e_1 = 0$ and $e_2 = 1$.

If $0 < \gamma > \frac{\lambda-1}{k_H} - 1$ then the external principal pays bonus 18 in the interval $[p^*, 1/k)H]$, the agents works $e_1 = 1$ and $\mathbf{E}e_2 = 1 - p$.

Functions f_2, h_2 are more complicated to analyze. To skip uninteresting cases I concentrate on the ones that provide sufficient conditions for the cases similar to those described in the Proposition 3. Note that the position of the maximum of f_2 is described in the Proposition 3.

One of the cases is such that the straight line h_2 lies above f_2. In this case, the external principal is eager to pay the bonus (22) in the first period for any $p \in [1/k_H, 1]$. Sufficient condition is $h_2(1) = \gamma > f_2^{\max}$ where $f_2^{\max}$ is defined in the Proposition 3. Then $e_1 = 1$ and only the agents with low costs work in the second period, $\mathbf{E}e_2 = 1 - p$.

Another case is one in which $h_2(1/k_H) = \gamma(2 - 1/k_H) > f_2(1/k_H)$, $h_2(1) = \gamma < f_2(1) = k_H - 1$. In this case, the external principal pays the bonus (22) for $p \in [1/k_H, p^\gamma_{R1}]$ and does not pay the bonus for $p \in (p^\gamma_{R1}, 1]$. Here p^γ_{R1} is defined from the equation

$$-p^2(\lambda-1) + p(\lambda + k_H) - 2 = \gamma(2-p), -p^2(\lambda-1) + p(\lambda + k_H + \gamma) - 2 - 2\gamma = 0,$$

$$p^\gamma_{R1} = \frac{\lambda + k_H + \gamma - \sqrt{(\lambda + k_H + \gamma)^2 - 4(\lambda-1)(2+2\gamma)}}{2(\lambda-1)}.$$

In this case for $p \in [1/k_H, p^\gamma_{R1}]$ we get $e_1 = 1$, $\mathbf{E}e_2 = 1 - p$, and for $p \in (p^\gamma_{R1}, 1]$ we get $e_1 = 0$, $e_2 = 0$.

The last important situation is one in which the bonus is always zero in the interval $[1/k_H, 1]$. This case takes place when $h_2(1/k_H) = \gamma(2 - 1/k_H) < f_2(1/k_H)$, $h_2(1) = \gamma < f_2(1) = k_H - 1$, e.g. for small γ. External principal always pays zero, so the agent's efforts $e_1 = e_2 = 0$.

Finally, note that if I define $C_1 = \min\left(\frac{k_H}{2k_H-1}\left(\frac{\lambda}{k_H} - 1 - \frac{\lambda-1}{k_H^2}\right), k_H - 1\right)$, then if $\gamma \le C_1$

there is no intersection between h_2 and f_2. Both principals will pay zero and the agent works 0 in both periods. If $g \geq C_1$ then either $h_2(1/k_H) > f_2(1/k_H)$, $h_2(1) > f_2(1)$, or both. Then either $h_2 > f_2$ in the whole interval, or they cross inside the interval. Thus at least in one interval inside $[1/k_H, 1]$ the external principal will pay bonus (22) and the agent will work $e_1 = 1$ and $\mathbf{E}e_2 = 1 - p$.

This concludes the proof of the Proposition 4.

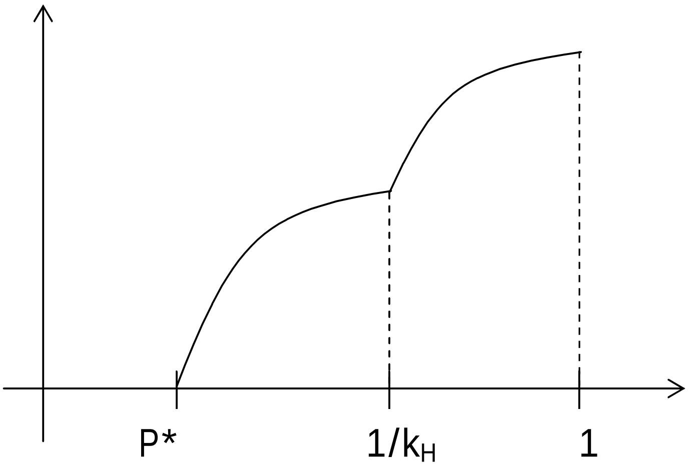

Figure 1: Cases A.1, B.1 and C.1.

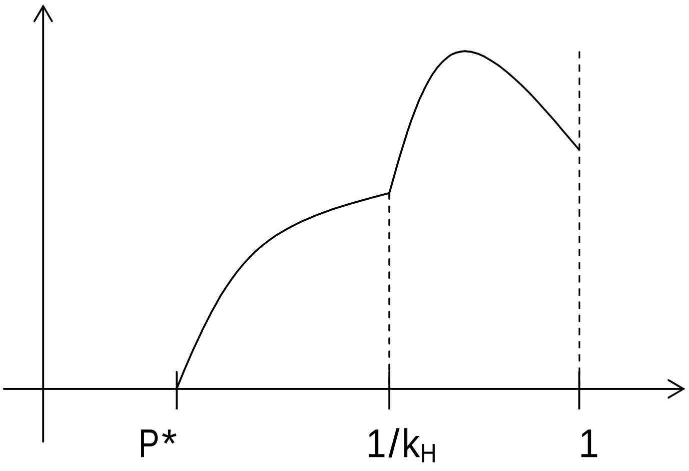

Figure 2: Cases B.2 and C.2

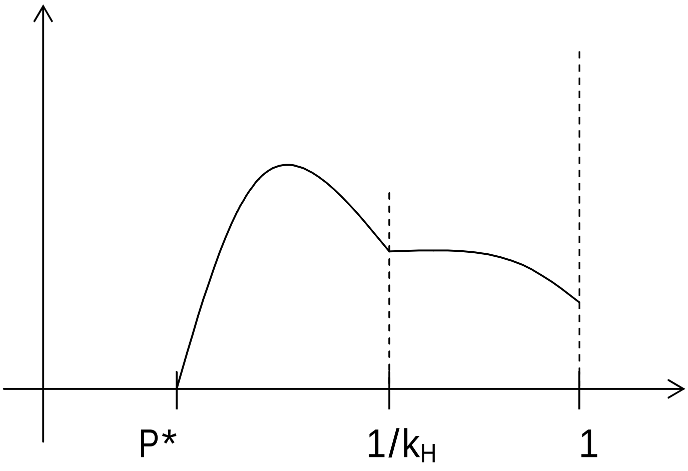

Figure 3: Cases A.4 and B.4

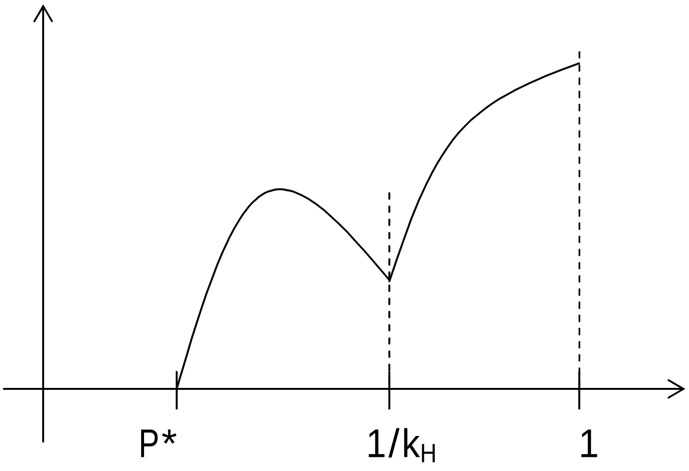

Figure 4: Case A.2

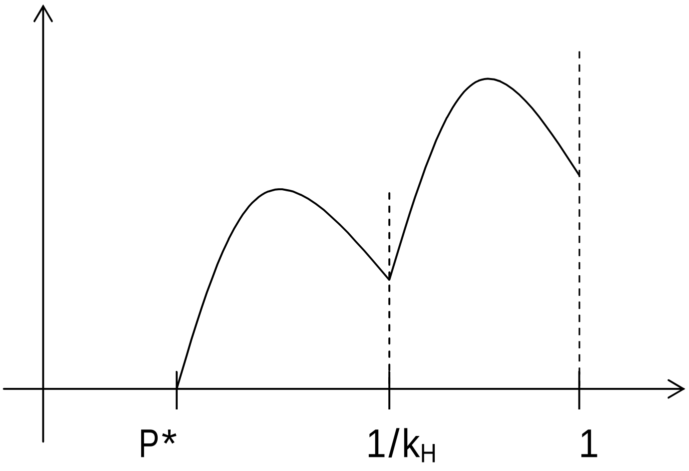

Figure 5: Cases A.3 and B.3