On Inducing Agents with Term Limits to Take Appropriate Risk*

Rohan Dutta† and Pierre-Yves Yanni‡

September 11, 2017

Abstract

A principal needs agents to appropriately select a risky action over a safe one, with replacement as the only incentivizing tool. The risky action can reveal the agent's decision-making competence. Aghion and Jackson (2017) shows how the inability to commit to a retention policy severely limits the principal's welfare. We find that adding a two-period term limit for agents helps the principal considerably. It allows the principal to enforce a random dismissal rule following the safe action. The incentive problem becomes even less severe if the agent has private information about his competence and the stakes are high.

JEL classification: D72, D82, D86, C72

Keywords: Term-limits, elections, principal-agent, discretion, replacement.

1   Introduction

Consider the problem of inducing an agent to choose, when appropriate, a risky action over a safe one, when dismissal is the only way the principal can incentivize. The action may reveal the agent's lack of competence, making his dismissal preferable to the principal. It turns out that the incentive problem is difficult to overcome if the principal cannot commit to keep the agent. We study this incentive provision problem in the presence of two-period term limits.

We find that two-period term limits can benefit the principal considerably. In particular, the term limit structure allows the principal to mimic the ability to make commitments and enforce a random dismissal rule if the agent chooses the safe action. This in turn incentivizes the agent to take the risky action appropriately more often. The incentive problem becomes even less severe if the agent knows about her own competence a lot better than the principal.

As observed in Holmstrom (1999), inducing the "appropriate" choice of risky actions can be as important as providing incentives for greater effort in the so-called managerial incentive problem. More generally, in a variety of economic environments, ranging from the choice of appropriate public policy to selecting the best research project to fund, it is not so much the level of effort that matters but the correct use of discretion.[1] The presence of risky actions, that bring a reward when appropriate and a loss when not, make competent agents who can perceive the appropriateness of said actions, desirable. The principal would prefer to take a chance on a new hire than retain an incompetent agent. An agent, uncertain of his own competence and wanting solely to remain in office, would then want to avoid actions that could reveal incompetence. Furthermore, as is typical in public office, the principal may be unable to vary monetary rewards, needing to rely solely on a credible threat of dismissal. This is the strategic environment studied in Aghion and Jackson (2017) (henceforth AJ).

AJ obtain the remarkable result that for low replacement costs if the principal cannot commit to a retention policy in advance and if there is no term limit, then in any Markov perfect equilibrium the principal can do no better than if she were to replace the agent every period. They then study whether the limited commitment ability implicit in term limits can help. In a cursory discussion of two-period term

---

[1]Consider, for instance, the task of drug approval by the FDA in the US.

limits they mention that for low replacement costs the principal would not do better than with a one-period term limit. By doing a careful and focused analysis of two-period term limits we find this to be inaccurate.

Our analysis deals squarely with agents facing two-period term limits. We obtain a rich set of results. Even with the commitment implicit in a two-period term limit, the highest equilibrium payoff achievable by the principal depends on whether the principal can further commit to a second term retention policy (before the first term) or not. We explicitly characterize these *optimal commitment retention policies* for different levels of replacement costs (Proposition 1). It turns out that committing to a one-period term limit is never optimal and for low replacement costs the optimal policy involves probabilistic replacement following the safe action. Without commitment ability, we always find the equilibrium alluded to in AJ, in which the agent always picks the safe action and is not replaced (Proposition 2). As AJ observe, for low replacement costs, this equilibrium indeed does worse for the principal than a one-period term limit. For low replacement costs, however, there exists an equilibrium that does significantly better for the principal. It features a random replacement rule following the safe action and the agent chooses the risky action appropriately more often (Proposition 3). *The construction relies on a key aspect of two-period term limits, the fact that an agent about whom the principal has learnt nothing from a first term may nevertheless be valued less than a replacement who in expectation is of equal competence.*

Continuing with two-period term limits and without commitment ability for the principal, we check if the agent having superior information about his own type helps the principal in equilibrium. The asymmetric information environment presents a technical difficulty in that the existence of an equilibrium, in which the principal's retention policy is solely driven by her belief and the first term action, is not guaranteed. Banks and Sundaram (1998) and Duggan (2017) deal with essentially the same problem, albeit in a different asymmetric information environment.[2] Indeed the most immediate candidate strategy profile, in which the principal never replaces a first term agent and the incompetent agent always plays the safe action, is an equilibrium only for high replacement costs (Proposition 4). Nevertheless, relying on a fixed point argument we establish that for lower replacement costs, an equilibrium satisfying our criteria does exist (Proposition 5) with intuitive comparative statics (Proposition 6).

---

[2]We discuss the relationship in greater detail in Section 2.3.

Finally we show that when the stakes associated with the risky action are comparable and high then the principal does even better in the asymmetric information setting than the symmetric information one (Proposition 7).

The rest of the paper is as follows. Section 2.1 describes the formal framework of the incentive problem. Section 2.2 deals with the setting of symmetric uncertainty, with commitment ability for the principal in 2.2.1 and without in 2.2.2. Section 2.3 deals with the asymmetric information setting. In section 3.1 we discuss how our infinite horizon model with two-period term limits compares to both a two-period model and the infinite horizon model without term limits. Finally in section 3.2 we discuss a few implications of our findings on the issue of incumbency advantage.

2   Incentives for Risk Taking

2.1   Framework

Ours is an infinite horizon principal agent model, in which time proceeds in discrete periods $t \in \{1, 2, \ldots\}$. In each period there are two equally probable states of the world, X, Y and two feasible actions x, y. The payoff in a given period to the principal are as below

	X	Y
x	0	0
y	$-d$	g

with

$$g - d < 0 \qquad and \qquad g, d > 0$$

The choice between x and y is made by the agent, who is either competent or incompetent. Agents live for two periods while the principal is infinitely-lived. Following an agent's first term (period), the principal must decide whether to keep or replace him at some cost, c (he is always replaced at the end of his second term).

A choice of y yields a success if the state of the world is Y and failure if it is X. If the agent chooses y the principal observes if it is a success or a failure. If instead he chooses x, the principal receives no further information about the state of the

4

world. The principal and agent discount the future with a factor, $\delta \in (0, 1)$. If the principal's payoff in period t is given by u_t then her total discounted payoff at period 1 is $\sum_{t=1}^{\infty} \delta^{t-1} u_t$. We use the variable W for the principal's total discounted payoff and assign appropriate superscripts to denote different cases. We use the notion of average discounted payoff when discussing the principal's welfare. Recall that for a given W the average discounted payoff is simply the constant utility that the principal needs in each period and in perpetuity to get a total discounted payoff of W. This is simply $(1 - \delta)W$.

Newly hired agents are competent with probability λ_0. In each term, the agent receives a signal about the state before he makes his choice. The signal is correct with probability $1/2$ (and therefore uninformative) if the agent is incompetent. A competent agent always receives the correct signal.[3]

The agent only cares about being in office. This can be modelled as the agent receiving some positive payoff in a given period if he is in office and 0 otherwise. The principal's preferences each period are as stated in the table above.

Notice that since $g > 0$ if a competent agent were to follow his signal Y by playing y the payoff to society would be positive. On the other hand, the assumption of $g - d < 0$ ensures that if an incompetent agent were to follow his signal Y by playing y, the payoff to society would be negative. We now make the following assumption.

$$\lambda_0 g + (1 - \lambda_0)\frac{g - d}{2} > 0,$$

so that it is better for society that a newly hired agent who does not know his type follows the signal.

Solution Concept: We study a refinement of perfect Bayesian equilibria of the principal agent game. Given the repeated game structure of our setting, the principal could sustain a high payoff in equilibria with history dependent strategies where the principal forces herself to take a suboptimal action today for fear of reverting to suboptimal equilibria tomorrow onwards. We do not find such equilibria plausible in our setting. The leading example of the principal both in our study and AJ is that of voters electing public officials. The ability of voters to self-discipline themselves using such history-dependent strategies, we believe, is limited.

---

[3]This assumption makes the analysis significantly cleaner. Our results continue to hold if the competent agent receives the correct signal with a sufficiently high probability.

We focus on perfect Bayesian equilibria in which both the principal and the agents use stationary strategies. In particular, the principal's choice between keeping and replacing an agent following a first term depends solely on her belief about his competence and the outcome of his first term. Likewise, all first term agents employ the same strategy. All second term agents use the same strategy which could depend upon the first term choices and outcomes.

Since the agent simply wants to remain in office, upon re-election into a second term, he is necessarily indifferent across all his actions. We focus on equilibria in which the agent in the second term does what is in the best interest of society. This allows us to focus on the key incentive problem which has to do with inducing a first term agent, motivated solely by re-election concerns, to take the risky action appropriately.

A few remarks are due. First, while we are studying a refinement of perfect Bayesian equilibria we are not imposing any restriction on the strategy space of either the principal or the agent. Indeed either could choose to deviate to complex history-dependent strategies. Second, it is by no means obvious, to the best of our knowledge, that such perfect Bayesian equilibria always exists. Indeed, showing their existence in our setting is a key contribution of our paper. Finally, perfect Bayesian equilibria that satisfy our criteria have the feature of anonymity in that all incumbents conditional on achieving the same outcome are treated alike. This makes transparent the way in which the two-period term limit breaks the symmetry between an incumbent and a challenger with identical expected competence.

In all that follows all mention of equilibrium correspond to perfect Bayesian equilibria that satisfy the aforementioned properties.

2.2   Symmetric Information

In this section we assume that the principal and the agent share the same uncertainty about the agent's type, captured by the prior, λ_0. The signal received by the agent is nevertheless private.

Consider the following first term strategy for the agent. He plays x upon observing X.[4] Let β be the probability that the agent plays y after observing the signal Y. Given this agent strategy, the principal Bayesian updates the type of the agent to

---

[4]This is without loss of generality for the purpose of our analysis, as we explain subsequently.

arrive at posteriors λ^s, λ^x and λ^f, upon observing a success, a choice of x and a failure, respectively:

$$\lambda^s = \frac{\lambda_0 \frac{1}{2}\beta}{\lambda_0 \frac{1}{2}\beta + (1-\lambda_0)\frac{1}{4}\beta} = \frac{\lambda_0}{\lambda_0 + (1-\lambda_0)\frac{1}{2}} = \frac{2\lambda_0}{1+\lambda_0},$$

$$\lambda^x = \frac{\lambda_0\left(1-\frac{1}{2}\beta\right)}{\lambda_0\left(1-\frac{1}{2}\beta\right) + (1-\lambda_0)\left(1-\frac{1}{2}\beta\right)} = \lambda_0,$$

$$\lambda^f = 0.$$

Let us define the principal's expected payoff from an agent's first term when his type is λ and his strategy is β:

$$u(\lambda, \beta) \equiv \frac{\beta}{2}\left[\lambda g + (1-\lambda)\frac{g-d}{2}\right].$$

It turns out that the optimal first term agent behaviour for society (and the principal) involves the agent following his signal. The benefit is two-fold. It generates the highest first term expected payoff for society. Additionally, with positive probability, it reveals if the agent is competent. This revelation allows the principal to replace an incompetent agent and retain a competent one. This possible revelation, of course, is what generates the incentive problem. A first term agent, caring solely about staying in office, may disregard his signal and choose the safe action x, thereby avoiding the possibility of being revealed incompetent and facing an ouster.

To induce the agent to follow his signal more closely and therefore take the risky action more appropriately, the principal needs to make the safe action less palatable or the risky action more attractive (or a combination of the two) for the agent. Her ability to do so depends greatly on whether or not she can commit to a second term retention policy before observing the first term outcomes. We consider the two cases in turn.

2.2.1 Commitment

The principal, in this section, can commit to a second term retention policy before the agent makes his first term choice. We continue to require that the decision to fire or retain must only be made conditional on the principal's belief about the agent's type

and the outcome of his first term. This boils down to a mechanism design problem in which the principal maximizes her infinite horizon discounted payoff by choosing an appropriate retention policy that applies to all agents at the end of their first term. We call this the optimal commitment retention policy.

It turns out that the optimal commitment retention policy induces the agent to follow his signal (i.e. playing x after observing X and y after Y) in the first term. Consider the following retention policies, which we refer to by their assigned labels.

- *Passive:* The agent is always retained.

- *Moderate:* The agent is replaced after a failure, retained after a success and retained with probability $\frac{1+\lambda_0}{2}$ after x.

- *Aggressive:* The agent is replaced after a failure, retained after a success and retained with probability $\frac{1-\lambda_0}{2}$ after x.

Each of these retention policies induce the agent to follow his signal. Nevertheless, they differ in how any information revealed in the first term is utilized. The passive policy does not directly use such information. However, it does allow the agent, by retaining him, to use this information about himself and make a better informed second term decision that best serves society. For instance, upon discovering that he is incompetent the agent would simply choose the safe action in the second term. The moderate and aggressive policies go further and with positive probability replace agents discovered to be incompetent.

To see where the specific numbers for the retention probability following x come from, consider the following. Suppose an agent is replaced after a failure and retained after a success, both with certainty. Then, after observing Y, the probability of obtaining a success (and being retained) upon playing y is $\lambda_0(1) + (1 - \lambda_0)(1/2) = (1 + \lambda_0)/2$. After observing X, the probability of the same is $\lambda_0(0) + (1 - \lambda_0)(1/2) = (1 - \lambda_0)/2$. Therefore, the probability of retaining an agent after x has to be between these two bounds to incentivize the agent to follow his signal.

Let
$$\tilde{c} = \frac{\delta}{4} \frac{(1 - \lambda_0)\lambda_0}{2} \frac{g + d}{2}.$$
We then have the following characterization.

8

Proposition 1. *The optimal commitment retention policy is*

$$Passive \ \ if \ c > \tilde{c} + \frac{2 + \delta + \delta\lambda_0}{4}u(\lambda_0, 1)$$

$$Moderate \ \ if \ \tilde{c} < c < \tilde{c} + \frac{2 + \delta + \delta\lambda_0}{4}u(\lambda_0, 1)$$

$$Aggressive \ \ if \ c < \tilde{c}$$

It should not be surprising that the threat of replacement cannot be profitably used in incentivizing appropriate behaviour if replacement is very costly. Indeed, we are mostly interested in settings where the replacement cost is small but nevertheless positive. Consider then, the average discounted payoff to society from the aggressive retention policy, which is optimal for $c < \tilde{c}$,

$$(1 - \delta)W^{Agg} = u(\lambda_0, 1) + \frac{\tilde{c} - c}{1 + \frac{\delta}{2}}.$$

It is instructive to compare this to the average discounted payoff to society from a one-term limit, which is never optimal,

$$(1 - \delta)W^{Oneterm} = u(\lambda_0, 1) - c.$$

2.2.2   No Commitment

Without commitment ability, for low replacement costs, the principal would end up replacing an agent who followed his signal in the first term and failed. At the same time, with positive replacement cost, the principal would rather retain an agent she has learnt nothing about, than replace him with an identical agent. Of course, this would make the agent choose to ignore his signal and simply take the safe action. This unraveling of incentives ensures the existence of a particularly inefficient equilibrium.

Proposition 2. *There is an equilibrium in which the agent always plays x in the first period and what is optimal for society in the second period. Society's average discounted payoff is $(1 - \delta)W = \frac{\delta u(\lambda_0, 1) - c}{1 + \delta}$.*

Notice that society does worse in this equilibrium than if it could commit to a one-term limit, an inefficient outcome already. The question is whether society can do any better. If the cost of replacement is high enough then society would

9

retain an agent irrespective of the first term outcome. The result is identical to what happens if the principal commits to never replace an agent. For smaller values of the replacement cost it would seem that we cannot do better than the inefficient outcome in Proposition 2. This is indeed the suggestion in AJ. It turns out, as the next proposition shows, we can do better.

Proposition 3. *If $c \leq \tilde{c}$, then the following is an equilibrium. The agent follows the signal Y with probability $\beta_1^* \equiv \frac{u(\lambda_0,1)+c}{u(\lambda_0,1)+\tilde{c}}$ and plays x otherwise and follows his signal in the second period. The principal replaces the agent after f, keeps him after s and retains him with probability $\frac{1+\lambda_0}{2}$ after x. Society's average discounted payoff is $(1-\delta)W = u(\lambda_0,1)$.*

The principal's ability to randomize without commitment ability rests on her indifference in equilibrium between keeping and replacing an agent who took the safe action in the first term. But recall that the principal's belief about the competence of an agent who takes the safe action is identical to that of a potential replacement. Given that replacement is costly, how could the principal be indifferent? This leads to the key insight of this model of term limits.

Remark 1. *In our construction a first term agent plays the risky action with positive probability thereby allowing the principal to learn his type and replace if need be. It is this option value of a first term agent that makes him more valuable than a second term agent who in expectation has the same competence. As a result, for an appropriate level of randomization by such a first term agent, the principal could in fact be indifferent, following a safe action, between allowing a second term and replacing with a higher valued first term agent at a cost.*

Consider the difference in average discounted payoff to the principal from the equilibria in propositions 3 and 2, $\frac{u(\lambda_0,1)+c}{1+\delta}$ where $c < \tilde{c}$. Recall that the payoff from proposition 2 is even worse than that from a one-term limit. The (average discounted) gain from the equilibrium in proposition 3 compared to the payoff from a one-term limit is c. This latter value, at first glance, may not seem substantial. Notice, however, that c could be as high as $\tilde{c}$. So, when $\tilde{c} = \frac{\delta}{4} \frac{(1-\lambda_0)\lambda_0}{2} \frac{g+d}{2}$ takes a high value, as (for instance) when the stakes (g and d) are high, the gain can be substantial. Finally, observe that none of the assumptions we have made impose any bound on how high $\frac{g+d}{2}$ can be.

2.3 Asymmetric Information

Holmstrom (1999) suggests that inadequate risk-taking behaviour in a standard signal-jamming model may be countered if the agent were better informed about his type than the principal. We verify this conjecture in our information setting.

While the principal continues to believe a new agent to be competent with probability λ_0, we now assume that the agent knows his own type. We leave all remaining assumptions about the timing of actions and preferences unchanged.

As before, we are interested in perfect Bayesian equilibria in which (a) the principal's choice between keeping and replacing an agent following a first term depends solely on his inferred competence following the outcome of his first term, (b) conditional on their type all first term agents behave alike and (c) the agent in the second term, given his type, does what is in the best interest of society. Furthermore, since we wish to verify if the introduction of such information asymmetry can improve the risk taking behaviour of agents, we focus on perfect Bayesian equilibria in which the competent type follows his signal. Of course, such an equilibrium may not exist. Indeed, a key contribution of our paper is to show that such an equilibrium does indeed exist.

Banks and Sundaram (1998) establish a similar existence result. Their framework entails the more standard mix of adverse selection and moral hazard. The relevant choice for the agent is akin to effort, the higher the better for the principal. Also standard are the agent types, with better types finding it cheaper to pick a given effort level. Duggan (2017) does a more complete analysis of a similar setting, while correcting an error in the existence proof of Banks and Sundaram (1998).

We denote by α the probability that the incompetent agent plays y (we do not condition the strategy on the signal because it is uninformative for him). If the competent agent follows his signal, Bayesian updating by the principal results in

$$\lambda^s(\alpha) = \frac{\lambda_0 \frac{1}{2}}{\lambda_0 \frac{1}{2} + (1 - \lambda_0)\frac{1}{2}\alpha} = \frac{1}{1 + \frac{(1-\lambda_0)}{\lambda_0}\alpha},$$

$$\lambda^x(\alpha) = \frac{\lambda_0 \frac{1}{2}}{\lambda_0 \frac{1}{2} + (1 - \lambda_0)(1 - \alpha)} = \frac{1}{1 + \frac{(1-\lambda_0)}{\lambda_0}2(1 - \alpha)},$$

$$\lambda^f(\alpha) = 0.$$

The principal's expected payoff from a given first term when her belief about the agent's type is λ and the incompetent type's strategy is α is

$$u^A(\lambda, \alpha) \equiv \frac{\lambda}{2}g + (1-\lambda)\alpha\frac{g-d}{2}.$$

Observe that following a choice of x, the principal's belief about the agent's type no longer remains constant. Whether it improves or gets worse depends on whether the probability with which the incompetent type plays the risky action is greater or less than $1/2$, the unconditional probability with which the risky action is appropriate. So, if the incompetent agent were to play the safe action with certainty in the first term, the principal upon observing the safe action would believe the agent to be competent with a lower probability than a replacement. While the replacement, if he turned out to be incompetent, would behave identically to an incompetent type in his second term, it is simply the higher probability of the replacement being competent that makes him more valuable than the incumbent to the principal. Such a replacement, therefore, would be profitable as long as the cost of replacement were not too high. This intuition is captured in the following proposition that shows how the incompetent agent choosing the safe action for sure can be supported in equilibrium if and only if the cost of replacement is sufficiently high.

The relevant replacement cost threshold is the following,

$$\hat{c} = (1+\delta)\frac{\lambda_0(1-\lambda_0)}{2-\lambda_0}\frac{g}{2}.$$

Proposition 4. *There exists an equilibrium in which the competent agent always follows his signal and the incompetent agent plays the safe action with certainty if and only if $c \geq \hat{c}$.*

In this equilibrium we can obtain a closed form solution for society's average discounted payoff, $(1-\delta)W = u^A(\lambda_0, 0) - \frac{c}{1+\delta}$, which is bounded above by $u^A(\lambda^x(0), 0)$.

As discussed above, when it is cheaper to replace the agent, the principal may be compelled to replace following a choice of the safe action in the first term. Duggan (2017) refers to this as the commitment problem of voters. In particular, the principal wishes an incompetent agent to always choose the safe action. But such a strategy would necessarily make the principal infer the incumbent's competence to be less probable than the challenger's, resulting in replacement. The principal's (voters)

12

inability to commit to a retention rule can thus frustrate a simple way to get the incompetent type to not select the risky action.

Given that this rules out a natural candidate for stationary equilibrium, the latter's existence is no longer assured. It turns out, however, that a stationary equilibrium does indeed exist and entails randomization on the part of both the incompetent agent and the principal.

Proposition 5. *When $c < \hat{c}$, there exists an equilibrium in which the competent agent always follows the signal, the incompetent one mixes with $\alpha^* \in (0,1)$ and the principal replaces the agent upon observing f and with probability $\frac{1}{2}$ after x. Society's average discounted payoff is $(1 - \delta)W(\alpha^*) = u^A(\lambda^x(\alpha^*), 0)$.*

The probability with which the incompetent type plays the risky action in the equilibrium above, α^*, solves the following equation.

$$\frac{1}{1 + \frac{1-\lambda_0}{\lambda_0} 2(1-\alpha)} \frac{g}{2} = \frac{u^A(\lambda_0, \alpha) - c + \delta\lambda_0 \frac{3}{4}\frac{g}{2}}{1 + \frac{2+\lambda_0}{4}\delta}.$$

Equilibrium changes as a result of increases in the replacement cost, the loss d, the gain g or the prior probability of competence λ_0 can be described by the following figures.

Figure 1: Increase in c, d

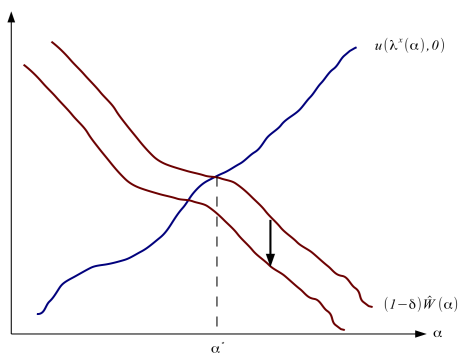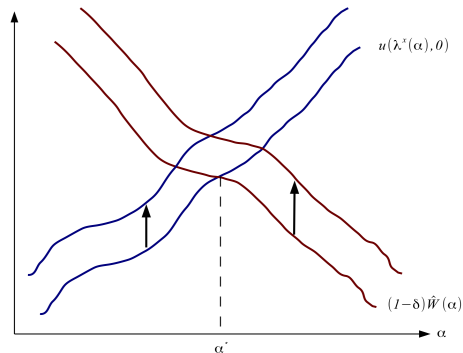Figure 2: Increase in g, λ

We collect these comparative statics in the following proposition.

Proposition 6. *An increase in the replacement cost (c) or the loss d, leads to a lower α^* and average discounted payoff, $(1 - \delta)W(\alpha^*)$. An increase in the gain g or the prior probability of competence, λ_0, leads to a higher average discounted payoff.*

13

Finally we compare the principal's payoff from the equilibria we have studied in the symmetric and asymmetric information settings. Let W^S and W^A denote the principal's infinite horizon discounted utility from the equilibrium in Proposition 3 and Proposition 5, respectively. We focus on $c < \min\{\hat{c}, \tilde{c}\}$ as required by these propositions.

Proposition 7. *If* $\frac{d-g}{g+d} < \frac{\lambda_0}{2}$ *then* $W^A > W^S$.

The condition requires the stakes, the gain and loss from the risky action, to be comparable and high. So, despite the possibility of excessive risk taking, the principal indeed benefits from the signalling incentive that the asymmetric information environment brings.[5]

3 Discussion

3.1 Two Period Model

Why do we not analyze a (seemingly simpler) two-period model instead of our infinite horizon model with two-period term limits? The two-period model is in fact closer to the infinite horizon model without term limits.[6] The intuition for this is simple. Return to the symmetric information setting and consider an agent who chooses the safe action in the first period. Neither he nor the principal has learnt anything further about his type. His expected level of competence is identical to that of the second period challenger. In a two-period model, there is no reason to expect that such an agent would behave any differently from a challenger in the second period. Both need to choose an action in this final period, which would have no further bearing on their payoff. In an infinite horizon model the situation is much the same. An agent who has learnt nothing about himself after his first period (having taken the safe action) is indistinguishable from a challenger. They both face identical future incentives. Of course, it makes little sense then to replace such an agent with the challenger at a cost.

By contrast, the key feature of term limits is the following. From the principal's perspective in a two-term limit infinite horizon model, an agent who has completed

---

[5]Chen (2015) shows that introducing asymmetric information in the Holmstrom (1999) two-period model leads to excessive risk taking.

[6]Duggan (2017) makes a similar observation apropos information settings featuring moral hazard and adverse selection. See also Duggan and Martinelli (2017).

one term can be fundamentally different from a challenger with two possible terms to serve, even if the latter's first term action revealed no information about him. This is because if first term agents pick the risky action with positive probability then the principal could learn his type and replace if incompetent. So even though the incumbent and the challenger are both equally competent in expected terms, more can be usefully learnt from the challenger. It is this asymmetry between the challenger and the incumbent that can be appropriately leveraged by the principal to provide the agent adequate incentives to take the appropriate risk.

3.2 Incumbency Advantage

Our model suggests that the issue of incumbency advantage depends on whether the uncertainty about the agent's competence is shared or one-sided. In the symmetric information setting, incumbency advantage does indeed exist. An incumbent who had taken the safe action, and is therefore of same expected competence as the challenger, is replaced with probability strictly less than $1/2$. In the asymmetric information setting the replacement probability following the safe action is exactly $1/2$. Note, however, that following a choice of the safe action in the asymmetric information setting, the incumbent's expected competence may be greater or less than the challenger's and in particular depends on the value of α^*.

A Appendix

Proof of Proposition 1. In what follows we use the following notation, stated formally below. $u(\lambda, X)$ and $u(\lambda, Y)$ capture the expected payoff to the principal in one period if an agent of type λ were to choose y following signals X and Y, respectively. λ_B^a is the updated probability assigned by the agent to his own competence after observing the result a of his first term action when the signal was B. For instance, λ_Y^s is the probability assigned by the agent to his own competence after observing his first term choice of y following a signal Y lead to success. A choice of x results in no learning and therefore, $\lambda_Y^x = \lambda_X^x = \lambda_0$. Finally $Pr(s|X) = 1 - Pr(f|X)$ is the probability with which a first term choice of y following a signal X leads to success. Similarly $Pr(s|Y) = 1 - Pr(f|Y)$ is the probability with which a first term choice of y following a signal Y leads to success.

The principal's problem is

$$W\left(\lambda_0\right) = -c +$$

$$
\max_{\substack{\beta_X,\beta_Y,\\p^s,p^x,p^f}}
\left\{
\begin{array}{l}
+\frac{\beta_Y}{2}\left(
\begin{array}{l}
u\left(\lambda_0,Y\right)\\[4pt]
+\Pr\left(s|Y\right)\left(
\begin{array}{l}
p^s\left(\delta\frac{\max\left\{u\left(\lambda_Y^s,Y\right),0\right\}+\max\left\{u\left(\lambda_Y^s,X\right),0\right\}}{2}+\delta^2 W\left(\lambda_0\right)\right)\\[4pt]
+\left(1-p^s\right)\delta W\left(\lambda_0\right)
\end{array}
\right)\\[10pt]
+\Pr\left(f|Y\right)\left(
\begin{array}{l}
p^f\left(\delta\frac{\max\left\{u\left(\lambda_Y^f,Y\right),0\right\}+\max\left\{u\left(\lambda_Y^f,X\right),0\right\}}{2}+\delta^2 W\left(\lambda_0\right)\right)\\[4pt]
+\left(1-p^f\right)\delta W\left(\lambda_0\right)
\end{array}
\right)
\end{array}
\right)\\[40pt]
+\frac{\beta_X}{2}\left(
\begin{array}{l}
u\left(\lambda_0,X\right)\\[4pt]
+\Pr\left(s|X\right)\left(
\begin{array}{l}
p^s\left(\delta\frac{\max\left\{u\left(\lambda_X^s,Y\right),0\right\}+\max\left\{u\left(\lambda_X^s,X\right),0\right\}}{2}+\delta^2 W\left(\lambda_0\right)\right)\\[4pt]
+\left(1-p^s\right)\delta W\left(\lambda_0\right)
\end{array}
\right)\\[10pt]
+\Pr\left(f|X\right)\left(
\begin{array}{l}
p^f\left(\delta\frac{\max\left\{u\left(\lambda_X^f,Y\right),0\right\}+\max\left\{u\left(\lambda_X^f,X\right),0\right\}}{2}+\delta^2 W\left(\lambda_0\right)\right)\\[4pt]
+\left(1-p^f\right)\delta W\left(\lambda_0\right)
\end{array}
\right)
\end{array}
\right)\\[40pt]
+\frac{1-\beta_Y}{2}\left(p^x\left(\delta\frac{\max\left\{u\left(\lambda_Y^x,Y\right),0\right\}+\max\left\{u\left(\lambda_Y^x,X\right),0\right\}}{2}+\delta^2 W\left(\lambda_0\right)\right)+\left(1-p^x\right)\delta W\left(\lambda_0\right)\right)\\[10pt]
+\frac{1-\beta_X}{2}\left(p^x\left(\delta\frac{\max\left\{u\left(\lambda_X^x,Y\right),0\right\}+\max\left\{u\left(\lambda_X^x,X\right),0\right\}}{2}+\delta^2 W\left(\lambda_0\right)\right)+\left(1-p^x\right)\delta W\left(\lambda_0\right)\right)
\end{array}
\right\}
$$

s.t.

$$
\begin{aligned}
\Pr\left(s|Y\right)p^s + \Pr\left(f|Y\right)p^f &\geq p^x \text{ if } \beta_Y > 0 && IC_Y\\
\Pr\left(s|X\right)p^s + \Pr\left(f|X\right)p^f &\geq p^x \text{ if } \beta_X > 0 && IC_X
\end{aligned}
$$

where

$$
\begin{aligned}
\Pr\left(s|Y\right) &= \lambda_0 + \left(1-\lambda_0\right)\frac{1}{2} = 1 - \Pr\left(f|Y\right)\\
\Pr\left(s|X\right) &= \left(1-\lambda_0\right)\frac{1}{2} = 1 - \Pr\left(f|X\right)
\end{aligned}
$$

$$
\begin{aligned}
u\left(\lambda,X\right) &= \lambda\left(-d\right) + \left(1-\lambda\right)\frac{g-d}{2}\\
u\left(\lambda,Y\right) &= \lambda g + \left(1-\lambda\right)\frac{g-d}{2}
\end{aligned}
$$

16

$$\lambda_Y^s = \frac{\lambda_0}{\lambda_0 + (1-\lambda_0)\frac{1}{2}} = \lambda_X^f > \lambda_0$$

$$\lambda_Y^f = 0 = \lambda_X^s < \lambda_0$$

$$\lambda_Y^x = \lambda_X^x = \lambda_0$$

We know that $u(\lambda_0, Y) > 0$ (by assumption) and $u(\lambda, X) < 0$ (because $d > g$) for all λ. Therefore

$$W(\lambda_0) = -c +$$

$$\max_{\substack{\beta_X, \beta_Y, \\ p^s, p^x, p^f}} \left\{ \begin{array}{l} +\frac{\beta_Y}{2} \left(\begin{array}{l} u(\lambda_0, Y) \\ + \Pr(s|Y)\left(p^s\left(\delta\frac{u(\lambda_Y^s,Y)}{2} + \delta^2 W(\lambda_0)\right) + (1-p^s)\delta W(\lambda_0)\right) \\ + (1 - \Pr(s|Y))\left(p^f\left(\delta\frac{\max\{u(\lambda_Y^f,Y),0\}}{2} + \delta^2 W(\lambda_0)\right) + (1-p^f)\delta W(\lambda_0)\right) \\ -p^x\left(\delta\frac{u(\lambda_0,Y)}{2} + \delta^2 W(\lambda_0)\right) - (1-p^x)\delta W(\lambda_0) \end{array} \right) \\ +\frac{\beta_X}{2} \left(\begin{array}{l} u(\lambda_0, X) \\ + \Pr(s|X)\left(p^s\left(\delta\frac{\max\{u(\lambda_Y^f,Y),0\}}{2} + \delta^2 W(\lambda_0)\right) + (1-p^s)\delta W(\lambda_0)\right) \\ + (1 - \Pr(s|X))\left(p^f\left(\delta\frac{u(\lambda_Y^s,Y)}{2} + \delta^2 W(\lambda_0)\right) + (1-p^f)\delta W(\lambda_0)\right) \\ -p^x\left(\delta\frac{u(\lambda_0,Y)}{2} + \delta^2 W(\lambda_0)\right) - (1-p^x)\delta W(\lambda_0) \end{array} \right) \\ +p^x\left(\delta\frac{u(\lambda_0,Y)}{2} + \delta^2 W(\lambda_0)\right) + (1-p^x)\delta W(\lambda_0) \end{array} \right\}.$$

Next, we show that the expression multiplied by β_X is negative, so that $\beta_X = 0$ is

optimal.

$$u\left(\lambda_0, X\right) + \Pr\left(s\mid X\right)p^s\left(\delta\frac{\max\left\{u\left(\lambda_X^s, Y\right), 0\right\}}{2} - (1-\delta)\,\delta W\left(\lambda_0\right)\right)$$

$$+ \left(1 - \Pr\left(s\mid X\right)\right)p^f\left(\delta\frac{u\left(\lambda_X^f, Y\right)}{2} - (1-\delta)\,\delta W\left(\lambda_0\right)\right)$$

$$- p^x\left(\left(\delta\frac{u\left(\lambda_0, Y\right)}{2} - (1-\delta)\,\delta W\left(\lambda_0\right)\right)\right)$$

$$= u\left(\lambda_0, X\right) + \delta\left(\Pr\left(s\mid X\right)p^s\frac{\max\left\{u\left(\lambda_X^s, Y\right), 0\right\}}{2} + \Pr\left(f\mid X\right)p^f\frac{u\left(\lambda_X^f, Y\right)}{2} - p^x\frac{u\left(\lambda_0, Y\right)}{2}\right)$$

$$+ \left(p^x - \Pr\left(s\mid X\right)p^s - \Pr\left(f\mid X\right)p^f\right)(1-\delta)\,\delta W\left(\lambda_0\right)$$

$$\leq u\left(\lambda_0, X\right) + \delta\left(\Pr\left(f\mid X\right)p^f\frac{u\left(\lambda_X^f, Y\right)}{2} - p^x\frac{u\left(\lambda_0, Y\right)}{2}\right)$$

$$< u\left(\lambda_0, X\right) + \Pr\left(f\mid X\right)\frac{u\left(\lambda_X^f, Y\right)}{2}$$

$$= -\lambda_0 d + (1-\lambda_0)\frac{g-d}{2} + \frac{1+\lambda_0}{4}\left(\frac{2\lambda_0}{1+\lambda_0}g + \frac{1-\lambda_0}{1+\lambda_0}\frac{g-d}{2}\right)$$

$$< 0$$

since $d > g$. The first (weak) inequality is due to (IC_X) and $u\left(\lambda_X^s, Y\right) = 0$. The second inequality results from $\delta < 1$ and setting $p^f = 1$ and $p^x = 0$ (the terms they multiply are all non-negative). We have

$$W\left(\lambda_0\right) = -c + \tag{1}$$

$$\max_{\substack{\beta_Y, p^s, \\ p^x, p^f}}\left\{ +\frac{\beta_Y}{2}\left(\begin{array}{l} u\left(\lambda_0, Y\right) \\ + \Pr\left(s\mid Y\right)\left(p^s\left(\delta\frac{u\left(\lambda_Y^s, Y\right)}{2} + \delta^2 W\left(\lambda_0\right)\right) + (1-p^s)\,\delta W\left(\lambda_0\right)\right) \\ + \Pr\left(f\mid Y\right)\left(p^f\left(\delta\frac{\max\left\{u\left(\lambda_Y^f, Y\right), 0\right\}}{2} + \delta^2 W\left(\lambda_0\right)\right) + \left(1-p^f\right)\,\delta W\left(\lambda_0\right)\right) \\ - p^x\left(\delta\frac{u\left(\lambda_0, Y\right)}{2} + \delta^2 W\left(\lambda_0\right)\right) - (1-p^x)\,\delta W\left(\lambda_0\right) \end{array}\right) \\ + p^x\left(\delta\frac{u\left(\lambda_0, Y\right)}{2} + \delta^2 W\left(\lambda_0\right)\right) + (1-p^x)\,\delta W\left(\lambda_0\right) \right\}.$$

To show that $\beta_Y = 1$ is optimal, we first demonstrate that for any p^x, the expression

in front of β_Y is strictly positive if $0 < \beta_Y < 1$. But this means that $0 < \beta_Y < 1$ cannot be optimal as it is always better to increase β_Y. So if at the optimum $\beta_Y > 0$ it must be $\beta_Y = 1$. Finally, when we characterize the principal's payoff from the optimal commitment retention policy with $\beta_Y = 1$ it is directly seen to be higher than that achieved with $\beta_Y = 0$.

$$
\begin{aligned}
&\max_{p^s,p^f}
\left\{
\begin{array}{l}
u\left(\lambda_0, Y\right) \\
+\Pr\left(s|\,Y\right)\left(p^s\left(\delta\frac{u\left(\lambda_Y^s, Y\right)}{2} + \delta^2 W\left(\lambda_0\right)\right) + \left(1 - p^s\right)\delta W\left(\lambda_0\right)\right) \\
+\Pr\left(f|\,Y\right)\left(p^f\left(\delta\frac{\max\left\{u\left(\lambda_Y^f, Y\right),0\right\}}{2} + \delta^2 W\left(\lambda_0\right)\right) + \left(1 - p^f\right)\delta W\left(\lambda_0\right)\right) \\
-p^x\left(\delta\frac{u\left(\lambda_0, Y\right)}{2} + \delta^2 W\left(\lambda_0\right)\right) - \left(1 - p^x\right)\delta W\left(\lambda_0\right)
\end{array}
\right\} \\
=\ &\max_{p^s,p^f}
\left\{
\begin{array}{l}
u\left(\lambda_0, Y\right) \\
+\frac{\delta}{2}\left(\Pr\left(s|\,Y\right)p^s u\left(\lambda_Y^s, Y\right) + \Pr\left(f|\,Y\right)p^f \max\left\{u\left(\lambda_Y^f, Y\right),0\right\} - p^x u\left(\lambda_0, Y\right)\right) \\
+\left(p^x - \Pr\left(s|\,Y\right)p^s - \Pr\left(f|\,Y\right)p^f\right)\delta\left(1 - \delta\right)W\left(\lambda_0\right)
\end{array}
\right\} \\
=\ &u\left(\lambda_0, Y\right) + \frac{\delta}{2}\max_{p^s,p^f}\left\{\Pr\left(s|\,Y\right)p^s u\left(\lambda_Y^s, Y\right) + \Pr\left(f|\,Y\right)p^f \max\left\{u\left(\lambda_Y^f, Y\right),0\right\} - p^x u\left(\lambda_0, Y\right)\right\} \\
=\ &u\left(\lambda_0, Y\right) + \frac{\delta}{2}\max_{p^s,p^f}\left(\Pr\left(s|\,Y\right)p^s u\left(\lambda_Y^s, Y\right) + \Pr\left(f|\,Y\right)p^f \max\left\{u\left(\lambda_Y^f, Y\right),0\right\} - p^x u\left(\lambda_0, Y\right)\right) \\
>\ &\left(1 - p^x\frac{\delta}{2}\right)u\left(\lambda_0, Y\right) > 0
\end{aligned}
$$

The second equality comes from requiring the IC_Y condition to hold with equality. The first inequality comes from $u\left(\lambda_Y^s, Y\right) > 0$ and $\max\left\{u\left(\lambda_Y^f, Y\right),0\right\} \geq 0$ (and the fact that we can always find p^s and p^f such that the IC_Y condition holds with equality).

We have already established that $\beta_X = 0$ and $\beta_Y = 1$. In order to show that $p^s = 1$, it suffices to notice that

$$
W\left(\lambda_0\right) < \frac{\frac{u\left(\lambda_Y^s, Y\right)}{2}}{1 - \delta}
$$

because $c \geq 0$ and $\lambda_Y^s > \lambda_0$. Then, it is clear from

$$
W\left(\lambda_0\right) = -c +
$$

$$
\max_{p^s, p^x, p^f} \left\{ +\frac{1}{2} \left(\begin{array}{c} u\left(\lambda_0, Y\right) \\ + \Pr\left(s \mid Y\right) \left(p^s \delta \left(\frac{u\left(\lambda_Y^s, Y\right)}{2} - \left(1 - \delta\right) W\left(\lambda_0\right) \right) + \delta W\left(\lambda_0\right) \right) \\ + \Pr\left(f \mid Y\right) \left(p^f \left(\delta \frac{\max\left\{u\left(\lambda_Y^f, Y\right), 0\right\}}{2} + \delta^2 W\left(\lambda_0\right) \right) + \left(1 - p^f\right) \delta W\left(\lambda_0\right) \right) \\ - p^x \left(\delta \frac{u\left(\lambda_0, Y\right)}{2} + \delta^2 W\left(\lambda_0\right) \right) - \left(1 - p^x\right) \delta W\left(\lambda_0\right) \\ + p^x \left(\delta \frac{u\left(\lambda_0, Y\right)}{2} + \delta^2 W\left(\lambda_0\right) \right) + \left(1 - p^x\right) \delta W\left(\lambda_0\right) \end{array} \right) \right\}
$$

s.t

$$
\Pr\left(s \mid Y\right) p^s + \Pr\left(f \mid Y\right) p^f \geq p^x \qquad IC_Y,
$$

that $p^s = 1$ is optimal since p^s multiplies a positive term and relaxes the IC_Y constraint.

Then we have,

$$
W\left(\lambda_0\right) = -c +
$$

$$
\max_{p^x, p^f} \left\{ +\frac{1}{2} \left(\begin{array}{c} u\left(\lambda_0, Y\right) \\ + \Pr\left(s \mid Y\right) \left(\delta \frac{u\left(\lambda_Y^s, Y\right)}{2} + \delta^2 W\left(\lambda_0\right) \right) \\ + \Pr\left(f \mid Y\right) \left(p^f \left(0 + \delta^2 W\left(\lambda_0\right) \right) + \left(1 - p^f\right) \delta W\left(\lambda_0\right) \right) \\ - p^x \left(\delta \frac{u\left(\lambda_0, Y\right)}{2} + \delta^2 W\left(\lambda_0\right) \right) - \left(1 - p^x\right) \delta W\left(\lambda_0\right) \\ + p^x \left(\delta \frac{u\left(\lambda_0, Y\right)}{2} + \delta^2 W\left(\lambda_0\right) \right) + \left(1 - p^x\right) \delta W\left(\lambda_0\right) \end{array} \right) \right\}.
$$

This expression is decreasing in p^f. Now since IC_Y must be satisfied, then in the optimal contract p^f must satisfy

$$
p^f = \max\left\{ \frac{p^x - Pr(s \mid Y)}{Pr(f \mid Y)}, 0 \right\}
$$

So if in the optimal contract $p^x > Pr(s \mid Y)$, then IC_Y must bind. Otherwise, $p^f = 0$. In general we have

$$
(1 - \delta) W(\lambda_0) \left\{ 1 + p^x \delta - \frac{\delta}{2} \left(p^x - \Pr\left(s \mid Y\right) p^s - \Pr\left(f \mid Y\right) p^f \right) \right\}
$$

20

$$= -c + \frac{1}{2}\left\{ u(\lambda_0, Y) + \delta Pr(s|Y)\frac{u(\lambda_Y^S, Y)}{2} + \delta p^x \frac{u(\lambda_0, Y)}{2}\right\}$$

Now if in the optimal contract $p^x > Pr(s|Y)$, then IC_Y binds. So it must be that

$$(1 - \delta)W(\lambda_0) = \frac{-c + \frac{1}{2}\left\{\frac{u(\lambda_0, Y)}{2} + \delta Pr(s|Y)\frac{u(\lambda_Y^S, Y)}{2}\right\}}{1 + p^x\delta} + \frac{u(\lambda_0, Y)}{4}$$

On the other hand if in the optimal contract $p^x \leq Pr(s|Y)$, then $p^f = 0$. So we get

$$(1 - \delta)W(\lambda_0) = \frac{u(\lambda_0, Y)}{2} + \frac{\frac{1}{2}\delta Pr(s|Y)\left[\frac{u(\lambda_Y^S, Y)}{2} - \frac{u(\lambda_0, Y)}{2}\right] - c}{1 + p^x\frac{\delta}{2} + \frac{\delta}{2}Pr(s|Y)}$$

Notice that if c is higher than $\frac{1}{2}\left\{\frac{u(\lambda_0, Y)}{2} + \delta Pr(s|Y)\frac{u(\lambda_Y^S, Y)}{2}\right\}$ then the optimal contract must set the highest possible p^x in other words 1. This would mean that $p^f = 1$ too. If c is smaller than that but larger than $\frac{1}{2}\delta Pr(s|Y)\left[\frac{u(\lambda_Y^S, Y)}{2} - \frac{u(\lambda_0, Y)}{2}\right]$ then the optimal contract must set $p^x = Pr(s|Y)$. If c is smaller than that too then the p^x must be as small as possible. Since we must satisfy the weak opposite of the IC_X inequality this must mean $p^x = Pr(s|X)$. □

Proof of Proposition 2. The present discounted payoff to the principal from this strategy profile at the start of a first term is

$$W = -c + 0 + \delta u(\lambda_0, 1) + \delta^2 W$$

which gives

$$(1 - \delta)W = \frac{\delta u(\lambda_0, 1) - c}{1 + \delta}.$$

First term agents have no incentive to deviate from choosing x since they are retained with probability 1. Following a first term choice of x, by retaining the agent as the strategy specifies the principal gets a present discounted payoff of $u(\lambda_0, 1) + \delta W$; replacing the agent brings $-c + W$. For replacement to be a profitable deviation it must be that $(1 - \delta)W > u(\lambda_0, 1) + c$, which is false since $(1 - \delta)W = \frac{\delta u(\lambda_0,1)-c}{1+\delta}$. □

Proof of Proposition 3. Given that he is replaced with probability $\frac{1-\lambda_0}{2}$ after x and fails with probability $\frac{1-\lambda_0}{2}$ when he follows the signal Y, the agent is indifferent be-

tween x and y when the signal is Y. Recall that $\beta_1^* \equiv \frac{u(\lambda_0,1)+c}{u(\lambda_0,1)+\tilde{c}}$, where $\tilde{c} = \frac{\delta}{4}\frac{(1-\lambda_0)\lambda_0}{2}\frac{g+d}{2}$. Given that the agent plays y with probability β_1^* after the signal Y, the principal's discounted payoff at the start of a first term is

$$
\begin{aligned}
W &= u(\lambda_0, \beta_1^*) - c + \left(\frac{1}{2} + \frac{1-\beta_1^*}{2}\right)\frac{1+\lambda_0}{2}\left(\delta u(\lambda_0, 1) + \delta^2 W\right) \\
&\quad + \frac{\beta_1^*}{2}\left(\lambda_0 + \frac{1-\lambda_0}{2}\right)\left(\delta u\left(\frac{2\lambda_0}{1+\lambda_0}, 1\right) + \delta^2 W\right) \\
&\quad + \left(\left(\frac{1}{2} + \frac{1-\beta_1^*}{2}\right)\frac{1-\lambda_0}{2} + \frac{\beta_1^*}{2}\frac{1-\lambda_0}{2}\right)\delta W \\
&= \beta_1^* u(\lambda_0, 1) - c + \left(1 - \frac{\beta_1^*}{2}\right)\frac{1+\lambda_0}{2}\delta u(\lambda_0, 1) \\
&\quad + \frac{\beta_1^*}{2}\frac{1+\lambda_0}{2}\delta u\left(\frac{2\lambda_0}{1+\lambda_0}, 1\right) + \frac{1-\lambda_0}{2}\delta W + \frac{1+\lambda_0}{2}\delta^2 W \\
&= \frac{\beta_1^* u(\lambda_0, 1) - c + \left(1 - \frac{\beta_1^*}{2}\right)\frac{1+\lambda_0}{2}\delta u(\lambda_0, 1) + \frac{\beta_1^*}{2}\frac{1+\lambda_0}{2}\delta u\left(\frac{2\lambda_0}{1+\lambda_0}, 1\right)}{(1-\delta)\left(1 + \frac{1+\lambda_0}{2}\delta\right)} \\
&= \frac{u(\lambda_0, 1)}{1-\delta}
\end{aligned}
$$

The principal is indifferent between replacing and keeping the agent after observing x in the first term since

$$
u(\lambda_0, 1) + \delta W = u(\lambda_0, 1) + \delta\frac{u(\lambda_0, 1)}{1-\delta} = \frac{u(\lambda_0, 1)}{1-\delta} = W.
$$

$\square$

Proof of Proposition 4. The principal's present discounted payoff from the stated strategy profile is

$$
W = (1+\delta)u^A(\lambda_0, 0) - c + \delta^2 W \Rightarrow W = \frac{(1+\delta)u^A(\lambda_0, 0) - c}{1 - \delta^2}.
$$

It is easy to see that the given the principal's strategy the agent of either type has no incentive to deviate.

Following a first term choice of x if the principal were to follow her strategy as stated she gets a present discounted payoff of $u^A(\lambda^x(0), 0) + \delta W$; instead, replacing the agent brings W. So for the principal to not have an incentive to deviate we must

have

$$u^A\left(\lambda^x\left(0\right),0\right)+\delta W \geq W$$

$$\Leftrightarrow \quad u^A\left(\lambda^x\left(0\right),0\right) \geq u^A\left(\lambda_0,0\right)-\frac{c}{1+\delta}$$

$$\Leftrightarrow \quad c \geq \left(1+\delta\right)\left(u^A\left(\lambda_0,0\right)-u^A\left(\lambda^x\left(0\right),0\right)\right)=\left(1+\delta\right)\frac{\lambda_0\left(1-\lambda_0\right)}{2-\lambda_0}\frac{g}{2}.$$

□

Proof of Proposition 5. Given the principal's strategy, it is optimal for the competent agent to follow her signal. When she observes Y, she obtains s and is kept with probability 1. When she observes X, she plays x and is kept with probability $\frac{1}{2}$ instead of being replaced with probability one if she plays y (and obtains f). The incompetent agent is indifferent between x and y since in both cases she is replaced with probability $\frac{1}{2}$.

Next, we show that there exists a probability $\alpha^* \in (0,1)$ for the incompetent agent to play y (which does not depend on the signal since it is uninformative for him) such that the principal is indifferent between keeping or replacing the agent after observing x. Denote by $\hat{W}$ the present discounted value for the principal if the players follow the strategy described above, with incompetent agent playing y with probability α. This value for the principal is given by

$$
\begin{aligned}
\hat{W}\left(\alpha\right) =\ & u\left(\lambda_0,\alpha\right)-c+\delta\lambda_0\frac{3}{4}\frac{g}{2} \\
& +\left[\left(\frac{\lambda_0}{2}+\left(1-\lambda_0\right)\left(1-\alpha\right)\right)\frac{1}{2}+\left(\frac{\lambda_0}{2}+\left(1-\lambda_0\right)\frac{\alpha}{2}\right)\right]\delta^2\hat{W}\left(\alpha\right) \\
& +\left[\left(\frac{\lambda_0}{2}+\left(1-\lambda_0\right)\left(1-\alpha\right)\right)\frac{1}{2}+\left(1-\lambda_0\right)\frac{\alpha}{2}\right]\delta\hat{W}\left(\alpha\right) \\
=\ & \frac{u\left(\lambda_0,\alpha\right)-c+\delta\lambda_0\frac{3}{4}\frac{g}{2}}{\left(1-\delta\right)\left(1+\frac{2+\lambda_0}{4}\delta\right)}.
\end{aligned}
\tag{2}
$$

Then

$$(1-\delta)\,\hat{W}\,(0) = \frac{u\,(\lambda_0, 0) - c + \delta\lambda_0\frac{3}{8}g}{1 + \frac{2+\lambda_0}{4}\delta}$$

$$= \frac{u\,(\lambda_0, 0) - c + \frac{3}{4}\delta u\,(\lambda_0, 0)}{1 + \frac{2+\lambda_0}{4}\delta}$$

$$> \frac{u\,(\lambda_0, 0) + \frac{3}{4}\delta u\,(\lambda_0, 0) - (1 + \delta)\,(u\,(\lambda_0, 0) - u\,(\lambda^x\,(0), 0))}{1 + \frac{2+\lambda_0}{4}\delta}$$

$$= \frac{\frac{-1}{4}\delta u\,(\lambda_0, 0) + (1 + \delta)\,u\,(\lambda^x\,(0), 0)}{1 + \frac{2+\lambda_0}{4}\delta}$$

$$= \frac{\frac{-1}{4}\delta\lambda_0\frac{1}{2}g + (1 + \delta)\,\frac{\lambda_0}{2-\lambda_0}\frac{1}{2}g}{1 + \frac{2+\lambda_0}{4}\delta}$$

$$= \frac{\lambda_0}{2 - \lambda_0}\frac{g}{2} = u\,(\lambda^x\,(0), 0)$$

since $c < (1 + \delta)\,(u\,(\lambda_0, 0) - u\,(\lambda^x\,(0), 0))$. Moreover,

$$(1-\delta)\hat{W}(2/3) = \frac{\frac{\lambda_0}{2}g + (1 - \lambda_0)\frac{2}{3}\frac{g-d}{2} - c + \delta\lambda_0\frac{3}{4}\frac{g}{2}}{1 + \frac{2+\lambda_0}{4}\delta} < \frac{\frac{\lambda_0}{2}g + \delta\lambda_0\frac{3}{4}\frac{g}{2}}{1 + \frac{2+\lambda_0}{4}\delta}$$

$$= \frac{\lambda_0 g}{2}\frac{1 + \frac{3}{4}\delta}{1 + \frac{2+\lambda_0}{4}\delta} \le \frac{\lambda_0 g}{2}\frac{3}{\lambda_0 + 2} = u(\lambda^x(2/3), 0).$$

Given that $u\,(\lambda^x\,(\alpha), 0)$ and $\hat{W}\,(\alpha)$ are continuous and respectively strictly increasing and decreasing in α, there exists a unique $\alpha^* \in (0, 2/3)$ that solves

$$u\,(\lambda^x\,(\alpha), 0) + \delta\hat{W}\,(\alpha) = \hat{W}\,(\alpha). \tag{3}$$

Since α^* satisfies Equation 3, the principal must be indifferent between retaining and replacing an agent after observing a first term choice of x, as required for our strategy profile to constitute an equilibrium. Since $\alpha^* < 2/3$ we know that $\lambda^x(\alpha^*) < \lambda^s(\alpha^*)$. This in turn ensures that following a successful first term voters optimally choose to retain the agent. Finally since $\frac{u(\lambda^x(\alpha^*), 0)}{1-\delta} = \hat{W}(\alpha^*) > 0$, the principal's choice to replace the agent after a failure is indeed optimal. $\square$

Proof of Proposition 7. To keep the two information settings distinct we adopt the following notation. The principal's expected payoff in a given period in the *symmetric information setting*, when the agent's expected type is λ and he plays x after X and

24

y with probability β upon observing Y, is given by

$$u^S(\lambda, \beta) \equiv \frac{\beta}{2}\left[\lambda g + (1-\lambda)\frac{g-d}{2}\right].$$

In the *asymmetric information setting* the principal's expected payoff in a given period, if the agent's competent type follows his signal and the incompetent type plays y with probability α, is given by

$$u^A(\lambda, \alpha) \equiv \frac{\lambda}{2}g + (1-\lambda)\alpha\frac{g-d}{2}$$

where λ is the agent's expected type. Let W^S denote the present discounted value to the principal from the equilibrium described in Proposition 3. Recall that $W^S = \frac{u^S(\lambda_0,1)}{1-\delta}$. Similarly let $W^A(\alpha^*)$ denote the present discounted value to the principal from the equilibrium described in Proposition 5. So, $W^A(\alpha^*) = \frac{u^A(\lambda^x(\alpha^*),0)}{1-\delta}$.

As per the proposition we focus on $c \leq \min\{\hat{c}, \tilde{c}\}$. Recall that

$$\lambda^x(\alpha) \equiv \frac{1}{1 + \frac{1-\lambda_0}{\lambda_0}2(1-\alpha)}.$$

Define $\bar{\alpha}$ such that $u^A\left(\lambda^x\left(\bar{\alpha}\right), 0\right) = u^S\left(\lambda_0, 1\right)$:

$$\frac{1}{1+\frac{1-\lambda_0}{\lambda_0}2\left(1-\bar{\alpha}\right)}\frac{g}{2} = \frac{1}{2}\left(\lambda_0 g + (1-\lambda_0)\frac{g-d}{2}\right)$$

$$\Rightarrow \bar{\alpha} = \frac{1}{2}\left(1 - \frac{d-g}{4u^S\left(\lambda_0, 1\right)}\right) = \frac{1}{2}\left(1 - \frac{d-g}{2\left(\lambda_0 g + (1-\lambda_0)\frac{g-d}{2}\right)}\right).$$

Our assumption of $\frac{d-g}{g+d} < \frac{\lambda_0}{2}$ ensures that $\bar{\alpha} \in (0, \frac{1}{2})$. Now, α^* from Proposition 5 is defined by $\hat{W}\left(\alpha^*\right) = u^A\left(\lambda^x\left(\alpha^*\right), 0\right) + \delta\hat{W}\left(\alpha^*\right)$ where $\hat{W}(\alpha)$ is defined in Equation 2.

We have

$$
\begin{aligned}
&(1-\delta)\,\hat{W}\left(\bar{\alpha}\right) - u^A\left(\lambda^x\left(\bar{\alpha}\right),0\right) \\
&= \frac{u^A\left(\lambda_0,\bar{\alpha}\right) - c + \delta\lambda_0\frac{3}{4}\frac{g}{2}}{1+\frac{2+\lambda_0}{4}\delta} - u^S\left(\lambda_0,1\right) \geq \frac{u^S\left(\lambda_0,1\right) - c + \delta\lambda_0\frac{3}{4}\frac{g}{2}}{1+\frac{2+\lambda_0}{4}\delta} - u^S\left(\lambda_0,1\right) \\
&= \frac{\frac{1}{2}\left(\lambda_0 g + (1-\lambda_0)\frac{g-d}{2}\right) - c + \delta\lambda_0\frac{3}{4}\frac{g}{2}}{1+\frac{2+\lambda_0}{4}\delta} - \frac{1}{2}\left(\lambda_0 g + (1-\lambda_0)\frac{g-d}{2}\right) \\
&\geq \frac{\frac{1}{2}\left(\lambda_0 g + (1-\lambda_0)\frac{g-d}{2}\right) - \frac{1}{8}\delta(1-\lambda_0)\lambda_0\frac{g+d}{2} + \delta\lambda_0\frac{3}{4}\frac{g}{2}}{1+\frac{2+\lambda_0}{4}\delta} - \frac{1}{2}\left(\lambda_0 g + (1-\lambda_0)\frac{g-d}{2}\right) \\
&= \frac{1}{2}\delta(1-\lambda_0)\frac{d-g}{2\delta + \delta\lambda_0 + 4} > 0.
\end{aligned}
$$

The first inequality follows from $\bar{\alpha} < \frac{1}{2}$. Since $u^A\left(\lambda^x\left(\alpha\right),0\right)$ increases in α and $W\left(\alpha\right)$ is decreasing it must be that $\alpha^* > \bar{\alpha}$. This finally gives us

$$
W^A\left(\alpha^*\right) = \frac{u^A\left(\lambda^x\left(\alpha^*\right),0\right)}{1-\delta} > \frac{u^A\left(\lambda^x\left(\bar{\alpha}\right),0\right)}{1-\delta} = \frac{u^S\left(\lambda_0,1\right)}{1-\delta} = W^S.
$$

$\square$

References

[1] Aghion, Philippe, and Matthew O. Jackson, 2016. Inducing Leaders to Take Risky Decisions: Dismissal, Tenure, and Term Limits. American Economic Journal: Microeconomics, 8(3): 1-38.

[2] Banks, Jeffrey S., and Rangarajan K. Sundaram, 1998. Optimal Retention in Agency Problems. Journal of Economic Theory, 82, 293-323.

[3] Chen, Ying, 2015. Career Concerns and Excessive Risk Taking. Journal of Economics and Management Strategy, 24 (1), 110-130.

[4] Homstrom, Bengt, 1999. Managerial Incentive Problems: A Dynamic Perspective. The Review of Economics Studies, 66, 169-182.

[5] Duggan, John, 2017. Term Limits and Bounds on Policy Responsiveness in Dynamic Elections. Journal of Economic Theory, 170, 426-463.

[6] Duggan, John, and Cesar Martinelli. The Political Economy of Dynamic Elections: Accountability, Commitment and Responsiveness. Journal of Economic Literature, forthcoming.