A Model of Selection and Policy Concerns

Álvaro Delgado-Vega* Benjamin Shaver

University of Chicago *University of Chicago*

April 15, 2025

Abstract

We study rank-order competition between a principal and an agent. The principal designs a policy by weighing the goals she shares with the agent and those that divide them. The policy's success depends on the principal's unknown quality and the agent's decision to help or sabotage. After the policy succeeds or fails, a committee allocates a prize to the winner, selecting the actor with the higher expected quality. In equilibrium, the agent's selection concerns lead him to help so that he avoids blame for failure or shares credit for success. The principal benefits from exploiting the agent's selection concerns to shift the policy in her favor while obtaining the agent's help, even if doing so lowers the probability the principal is selected. This implies the principal's expected utility is maximized when paired with a mediocre agent who replaces her after a resounding failure.

JEL Codes: D73, D23, D72

Keywords: Organizations; Political economy; Career concerns; Leadership; Rank-Order Competition

*Álvaro Delgado-Vega: adelgadovega@uchicago.edu; Benjamin Shaver: blshaver@uchicago.edu. We are deeply grateful to Wioletta Dziuda, Antoine Loeper, and Johannes Schneider for their guidance and advice. This paper benefited from helpful discussions with Scott Ashworth, Benjamin Blumenthal, Ethan Bueno de Mesquita, Georgy Egorov, Justin Fox, Scott Gehlbach, Barton Lee, Zhaotian Luo, Antoni-Ítalo Moragas, Marc Möller, Spencer Pantoja, Konstantin Sonin, and Rubén Veiga. We also thank audiences at the University of Bern, University of Chicago, Washington University in St. Louis, MPSA, and SIOE. During part of this project, Á. Delgado-Vega was supported by a Postdoctoral Fellowship at the Chair of Political Economy and eDemocracy at ETH Zurich and before by the Doctoral Scholarship PRE2019-089897 at Universidad Carlos III de Madrid; he is grateful for his time there.

1 Introduction

In 1797, Napoleon Bonaparte—still a general—lobbied the current governors of France, the Directory, to authorize an expedition to Egypt. The value of conquering Egypt was obvious to the Directors, giving them a strong "policy" incentive to provide Napoleon with the necessary troops and equipment to launch the expedition.[1] But what were their political considerations? Having just returned victorious from his Italian campaign, Napoleon was a growing threat to the unpopular Directory. His success in Italy had catapulted his reputation, partly because it had been achieved despite the meager funding and reinforcements provided by the Directory. If he now triumphed in Egypt, who would get the credit? And more importantly, if he failed, who would bear the blame? Rather than creating an incentive for the Directors to sabotage Napoleon, these questions surrounding credit and blame strengthened their incentives to support his expedition. As Roberts (2014) explains:

> It was in the Directors' interests for Napoleon to go to Egypt. He might conquer it for France or—just as welcome—return after a defeat with his reputation satisfyingly tarnished.

The elements highlighted in this example are common to many organizations. Often, the person who devises a project is not the same person who puts the resources, time, or effort into implementing it. The "ideas person" and "implementation person" are frequently also competitors. A C.E.O. might propose a strategy for her firm that must be implemented by a senior executive who is perceived to be her successor. The mayor of a large city might propose a policy that is partially funded by the state governor, who fears the mayor could someday be his rival.

We study a model of rank-order competition between a principal (she) and her agent (he). The principal begins by choosing a project or goal for the organization, which we refer to as a "policy." The agent publicly chooses between helping execute the policy and sabotaging it. Whether the policy succeeds depends on the principal's unknown quality and the agent's decision. After observing the policy outcome, a committee updates its beliefs about the principal's quality. It then allocates a prize between the principal and agent—e.g., who leads the organization in the future—selecting the actor with the highest expected quality. Importantly, we assume failure is more informative about the principal's quality when it

[1] "Bonaparte's role in the inception of the expedition was fundamental" (Dwyer, 2007). He was also deeply involved in the planning of the expedition: "The directors asked Napoleon to estimate the requirements in men, munitions and money [...] Napoleon replied in early March, in effect stating his own terms. He called for an expeditionary force of nearly 25,000 infantry, 3,000 horses and 1,500 artillerymen. [...] The Directory accepted these requirements and in mid March put Napoleon in titular command of the expedition." (Asprey, 2008).

occurs despite the agent's help, and conversely, success is more informative when it occurs despite the agent's sabotage. The principal and agent have policy preferences—they care both about the content of the policy and whether it is successfully executed—and selection concerns—they want to be selected by the committee.

We first consider the case where, like in our opening example, the principal and agent have the same policy preferences. Our first result shows that while the agent's selection concerns can incentivize him to sabotage the principal, they can also incentivize him to help. The principal's and agent's expected qualities determine whether we are in the former or latter case. If the expected qualities are similar, the committee selects the agent after any failure, including failure partly due to sabotage. So sabotage is the agent's optimal path. However, this logic changes if the expected qualities are neither too close nor too far. Then, *how* success or failure occurs matters. For instance, when the committee only selects the agent if the principal fails 'big'—that is, despite the agent's help—the agent's selection concerns incentivize him to help so that he shifts the blame for failure to the principal. Similarly, when the committee only selects the principal if she succeeds 'alone'—that is, despite the agent's sabotage—the agent's selection concerns incentivize him to help so that he shares credit for success.

That competition incentivizes help reverses a common result in the literature on rank-order competition. Previous work finds that an increase in the prize spread—the difference in payoffs between the winner and loser of the competition—increases the incentive to sabotage (Lazear, 1989; Dye, 1984; Falk et al., 2008). This comparative static may be reversed in our model because the agent's decision affects the information revealed by successful or failed policy execution. We show that increasing the prize spread—captured in our model by the payoff from being selected by the committee—can increase the agent's incentive to help the principal.

The insight that the agent's selection concerns can incentivize him to help the principal extends to the case where the principal and agent have different policy preferences. For example, suppose the principal is an ambitious, politically progressive mayor of a large city who proposes a policy that would require financial support from the more moderate governor of the state. Our previous result shows that if the governor fears the mayor might challenge him in the future, he may have an incentive not to sabotage but to fund the project. Can the mayor take advantage of this incentive when initially designing the policy? We analyze the model when the principal and agent have misaligned policy preferences and show this is indeed the case. Specifically, we assume that although the policy has common interest aspects, it also has aspects that divide the principal and agent. At the start of the game, the principal sets the organization's agenda by weighing these aspects.

The principal faces a fundamental trade-off: either she makes a policy concession large enough to obtain the agent's support, or she pursues her preferred agenda, which the agent sabotages because he dislikes it. There is additional nuance to this trade-off in the region where the agent's selection concerns incentivize him to help, i.e., when the principal and agent's expected qualities are neither too close nor too far. On the one hand, the principal can leverage the agent's selection concerns to obtain his help on a policy the principal likes more. On the other hand, enlisting the agent's help increases the probability that the committee selects the agent. For example, a principal who the committee selects unless she fails 'big'—i.e., with the agent's help—can obtain help despite making a small concession. However, doing so opens the possibility that the committee selects the agent. In contrast, choosing her preferred agenda, which the agent sabotages, ensures the committee will select her.

Our second result shows that indeed the principal's incentive to obtain the agent's help is strongest in the region where she can leverage the agent's selection concerns, even though doing so means she is less likely to be selected. The benefit the principal receives from obtaining the agent's help on a policy she likes more is larger than the cost she pays in terms of a decreased probability of being selected. This result arises from a key distinction between how policy concessions and selection concerns affect the agent's incentive to help. A policy concession, once granted, cannot be reversed if the agent does not help. In contrast, the agent's benefit from an increased probability of selection is only realized if he helps, thereby influencing how the principal's quality is assessed. In other words, selection concerns create a benefit contingent on the agent's help, making it less costly for the principal to incentivize his help than solely using policy concessions.

Lastly, we analyze the principal's expected utility when paired with agents of different expected qualities. The principal's expected utility is lowest when the agent's expected quality is close to her own since any failure results in the committee selecting the agent, which leads the agent to sabotage. However, the principal's expected utility is not maximized when the agent is perceived as vastly inferior, which ensures the principal is selected by the committee in any circumstance. The reason is that in this scenario, the agent has no selection concerns for the principal to exploit. Indeed, the principal attains her highest expected utility when paired with "mediocre" agents—those who the committee only selects in the event of a resounding failure—as these agents have selection concerns for the principal to leverage.

2 Related Literature

The literature that studies the relationship between leaders and their collaborators using a principal-agent framework is wide and identifies tensions that arise due to various reasons like moral hazard, asymmetric information, or—like in our model—career concerns.[2] There are many reasons why career concerns may create problems in an organization, e.g., because the principal's promises lack credibility (Dewan and Myatt, 2007, 2010) or because high-quality collaborators have better outside options (Mattozzi and Merlo, 2008; Zakharov, 2016; Dessein and Garicano, 2023). In our paper, career concerns create tension in a context where the principal and agent are engaged in rank-order competition, e.g., because they compete to be the next leader of the organization.

In this vein, the two closest papers to ours are Zhou (2023), which studies the "crown-prince problem" in autocracies, and Geelen and Hajda (2024), which studies C.E.O. succession. Both papers identify why a leader might sabotage her heir to reduce the threat he poses to her leadership. Similar to these papers, we show that selection (or succession) concerns can induce disarray within the organization. However, we also identify—as a result of microfounding political capital as a belief about the agents' quality—that selection concerns can foster unity.

Our paper is also related to the literature on sabotage in rank-order competition (e.g., Falk et al. (2008); Drago and Garvey (1998); Dye (1984); Chen (2003)). The seminal paper on this topic, Lazear (1989), studies a setting where an agent wins a competition if his output—a function of his private effort and his competitor's private sabotage—exceeds his competitor's. In this setting, increasing the prize spread—the gap between the winner and loser's prizes—induces agents to shift from exerting effort to sabotaging the other player. We study a setting with incomplete information about the agents' types and a publicly observed decision to help or sabotage. Our results offer a different picture: increasing the prize spread may increase the agent's incentive to help since the agent helps to share credit for success and avoid blame for failure.

Finally, this paper is related to the literature on sabotage in politics. In Gieczewski and Li (2022), Hirsch and Kastellec (2022), and Kang (2022) parties sabotage the policies of their adversaries; in Heo and Wirsching (2024), bureaucrats sabotage policies they dislike. In these papers, as well as in ours, sabotage affects the information that is learned from failed policy execution. In Hirsch and Kastellec (2022) and Kang (2022), the papers closest to ours, the policy is determined exogenously. In contrast, we study a model where the principal endogenously chooses the policy, anticipating the threat of sabotage and the benefit of help. This allows us to study the interaction between sabotage and agenda-setting.

[2] With respect to asymmetric information, Prendergast (1993) shows that when a collaborator enjoys an information advantage relative to the leader, he may not share his private information because he is under the principal's subjective evaluation. In Egorov and Sonin (2011), a collaborator's information advantage translates into him 'back-stabbing' the leader. Blanes I Vidal and Möller (2007) study the reverse situation, where the principal has an information advantage relative to her collaborators and shows she is subject to motivational bias when sharing her information with them, which distorts information revelation.

3 Model

An organization has a committee and two actors: a principal (P, she) and an agent (A, he). Each actor $i \in \{P, A\}$ has an unknown type $\theta_i \in \{0, 1\}$. This type is either high ($\theta_i = 1$) or low ($\theta_i = 0$). Let $\pi_i \in (0, 1)$ denote the *ex ante* probability that i is the high type.

The organization executes a policy g, which either succeeds ($g = \overline{g}$) or fails ($g = \underline{g}$). The success of the policy is stochastic and depends on the principal's type and whether the agent helps ($h = 1$) or sabotages ($h = 0$). We denote the probability of successful execution as $f(\theta_P, h)$ and assume $f(\theta_P, h) \in [0, 1]$.

Timing. The game has three stages:

1. *Agenda-Setting Stage:* The principal publicly chooses a policy $x \in [0, \overline{x}]$.
2. *Policy Stage:* The agent publicly chooses whether to help execute or sabotage the policy, $h \in \{0, 1\}$. The policy succeeds with probability $f(\theta_P, h)$.
3. *Selection Stage:* The committee observes the agent's choice and the policy outcome and selects the principal or agent to receive a prize, e.g., to be the future leader of the organization, $c \in \{P, A\}$.

Information Structure. We impose the following structure on the probability of success $f(\theta_P, h)$: (*i.*) Success is more likely when the principal is the high type and when the agent helps, i.e., $f(\theta_P, h)$ is strictly increasing in θ_P and h. (*ii.*) Failure is more informative about the principal's type when it occurs with the agent's help:

$$\frac{1 - f(0, h)}{1 - f(1, h)} \text{ is weakly increasing in } h;$$

and success is more informative about the principal's type when it occurs despite the agent's sabotage:

$$\frac{f(1, h)}{f(0, h)} \text{ is weakly decreasing in } h.$$

The implication of (*ii.*) is that, irrespective of whether the policy succeeds or fails, the agent's help lowers the assessment of the principal's type. In the Supplementary Appendix, we show that (*ii.*) is satisfied as long as $f(\theta_P, h)$ does not feature too much complementarity between the principal's type and the agent's help.[3]

As an example, a natural functional form that satisfies these conditions is:

$$f(\theta_P, h) = v + mh + (r_0 + r_1 h)\theta_P, \tag{1}$$

where $v, r_1 \geq 0$, $m, r_0 > 0$, $v + m + r_0 + r_1 \leq 1$, and the complementarity between quality

[3] Specifically, we show that (*ii.*) is satisfied as long as $\frac{\partial^2 \ln(f(\theta_P,h))}{\partial\theta\partial h} \leq 0$. This condition is satisfied as long as $\frac{\partial^2 f(\theta_P,h)}{\partial\theta\partial h}$ is not too large.

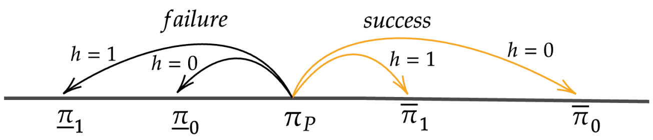

Figure 1: Updating on the principal's type.

and help is bounded, $\frac{mr_0}{v} \geq r_1$.[4] In particular, these conditions allow for the case of perfect substitution, i.e., $r_1 = 0$.

Let us define the *ex post* belief about the principal's type as $\overline{\pi}_h = \Pr(\theta_P = 1|g = \overline{g}, h)$ and $\underline{\pi}_h = \Pr(\theta_P = 1|g = \underline{g}, h)$.[5] Then, the assumed structure on $f(\theta_P, h)$ implies:

$$\underline{\pi}_1 \leq \underline{\pi}_0 < \pi_P < \overline{\pi}_1 \leq \overline{\pi}_0. \tag{2}$$

Figure 1 depicts this. Given belief π_P, we denote the expected probability the policy is successfully executed as:

$$\boldsymbol{f}(\pi_P, h) = \pi_P f(1, h) + (1 - \pi_P) f(0, h).$$

Payoffs. Each actor—the principal and the agent—has policy and selection concerns.[6] Actor i's policy concerns are captured by $p_i(g, x)$. W.l.o.g., we normalize the payoff from failed execution to zero: $p_P(g = \underline{g}, x) = p_A(g = \underline{g}, x) = 0$. For a given x, the payoffs from successful policy execution are:

$$p_P(g = \overline{g}, x) = \varphi + x,$$
$$p_A(g = \overline{g}, x) = \varphi - x,$$

where $\varphi \in \mathbb{R}$ and captures common interest aspects of the policy and $x \in [0, \bar{x}]$ and captures divisive aspects. The parameter $\bar{x}$ measures the extent of disagreement between the principal and agent, whose bliss points are $x = \bar{x}$ and $x = 0$, respectively.[7] The selection concerns come from the prize, $b > 0$ that actor i receives if selected by the committee. This prize could represent an office rent the leader of the organization enjoys.

[4] See the Supplemental Appendix for the derivation of these assumptions.

[5] In the Appendix, we derive the closed form version of these posterior beliefs and show they satisfy (2).

[6] Other papers assume that help or sabotage are costly actions (e.g, Lazear, 1989; Chen, 2003; Drago and Garvey, 1998). For simplicity, we treat both actions as costless. However, our qualitative results remain if either action carries a cost.

[7] Our results generalize qualitatively to a setting where, when the policy succeeds, the principal receives $\varphi - \ell_P(\bar{x}, x)$ and the agent, $\varphi - \ell_A(0, x)$, where ℓ_A and ℓ_P are concave, continuous loss functions that are increasing in the distance between x and the agent and principal's bliss points, respectively.

Actor i's utility function is:

$$u_i(g, x, c) = p_i(g, x) + \mathbb{1}_{c=i} b. \tag{3}$$

To focus on the most interesting regions of the parameter space, we assume that for any x, the prize is larger than the agent's policy payoff:

Assumption 1 *$b > \varphi$.*

Equilibrium. Our equilibrium concept is perfect Bayesian equilibrium (henceforth, "equilibrium"). A PBE requires that (*i.*) each player's choice is sequentially rational given their belief at the time of their action and the other players' strategies and (*ii.*) each player's belief about the principal's type satisfies Bayes' rule on the equilibrium path.

Discussion of the Model. Our model analyzes rank-order competition between a principal who plays the role of the "ideas person," and an agent who acts on her behalf, either helping or sabotaging the principal's policy. This type of competition might arise if the principal is the current leader of an organization and the agent is her presumptive heir. For instance, the principal could be the C.E.O. of a firm who proposes a strategy that needs to be executed by a senior executive. If the strategy fails, the board of directors—the committee—might remove the C.E.O. and promote the senior executive.[8] This model also applies to settings where the two actors are competitors. For example, the ambitious mayor of a large city may design a project that requires funding from the state governor. Apart from any policy disagreement, the governor might fear that successful implementation of the policy will boost the mayor's reputation with voters, positioning the major well to run against him.[9]

We assume the actors' contributions to the success of the policy are different. The principal's contribution is the quality of the design of the idea, which is related to her type. Was the Egyptian expedition proof of Napoleon's abilities as commander and organizer? Was the C.E.O.'s strategy cleverly designed? The agent's contribution is to help or sabotage policy execution, which is unrelated to her type.[10] The Directory chose to provide Napoleon with the troops and equipment he requested, but they could have provided him with less or even refused to authorize the expedition. The senior executive can execute the C.E.O.'s strategy correctly or shirk and sabotage. We also assume the actors' contributions to the

[8] One such example is Robert Nardelli and Frank Blake. Nardelli was the C.E.O. of Home Depot until he was pushed out, haunted by criticism of the slow growth of the company relative to its competitors (Barbaro, 2007; Kavilanz, 2007; Grow, 2007). He was replaced by Blake, his longtime protege, who had "played a key role in executing Nardelli's strategy at the retail chain" (Grow, 2007).

[9] For example, Andrew Cuomo, the governor of New York, and Bill de Blasio, the mayor of New York City, famously had an acrimonious relationship that affected the funding of programs de Blasio pursued in New York City like Pre-K.

[10] In the Supplementary Appendix, we show that our qualitative results remain, albeit in an altered region of the parameter space, if the agent's type affects the probability of success.

policy's success have either a limited degree of complementarity or are perfect substitutes. For instance, in the case of the mayor seeking funding for her project, abundant funding can partly compensate for pitfalls in its design.

4 Aligned Policy Preferences ($\bar{x} = 0$)

We begin our analysis by studying the case where the principal and agent have fully aligned policy preferences, i.e., $\bar{x} = 0$. This is the case of Napoleon's expedition to Egypt, where conquering Egypt was a strategic objective for the French in their war against Britain. It is also the case of a C.E.O. and a senior executive, both aiming to maximize their firm's profit.

The committee selects the agent if, after observing the policy outcome, it believes the agent is more likely to be a high type than the principal. Figure 2 shows that the committee's decision may be determined not only by whether the policy succeeds or fails but also by *how* that outcome occurs; that is, whether it occurs with the agent's help or despite his sabotage.

Lemma 1 *(i.) If π_P and π_A are far enough, i.e., $\pi_A < \underline{\pi}_1$ or $\overline{\pi}_0 < \pi_A$, the committee's selection is independent of the policy outcome:*

(a) If $\pi_A < \underline{\pi}_1$, the committee always selects the principal;

(b) If $\overline{\pi}_0 < \pi_A$, the committee never selects the principal.

(ii.) If π_P and π_A are close enough, i.e., $\underline{\pi}_0 < \pi_A < \overline{\pi}_1$, the committee's selection depends on the policy outcome: the committee selects the principal if and only if the policy succeeds.

(iii.) Otherwise, π_P and π_A are in an intermediate region, i.e., $\underline{\pi}_1 < \pi_A < \underline{\pi}_0$ or $\overline{\pi}_1 < \pi_A < \overline{\pi}_0$, where the committee's selection depends on the policy outcome and the agent's choice:

(a) If $\underline{\pi}_1 < \pi_A < \underline{\pi}_0$, the committee selects the principal unless the policy fails despite the agent's help;

(b) If $\overline{\pi}_1 < \pi_A < \overline{\pi}_0$, the committee only selects the principal if the agent sabotages and the policy succeeds.

This lemma allows us to differentiate when the agent's selection concerns incentivize him to help and when they incentivize him to sabotage. In case (*i.*), there are no selection concerns since the committee's selection is unrelated to the policy outcome. The agent's decision to help or sabotage depends entirely on his policy concerns: he helps if he wants the policy to succeed, i.e., $\varphi \geq 0$.

In comparison, in case (*ii.*), the committee's selection depends entirely on the policy outcome: failure—even if partly due to the agent's sabotage—leads to the agent's selection, while success—even if partly due to the agent's help—leads to the principal's selection. If $\varphi > 0$, he faces a trade-off between helping, which increases the probability of receiving the policy payoff φ, and sabotaging, which increases his probability of being selected by making

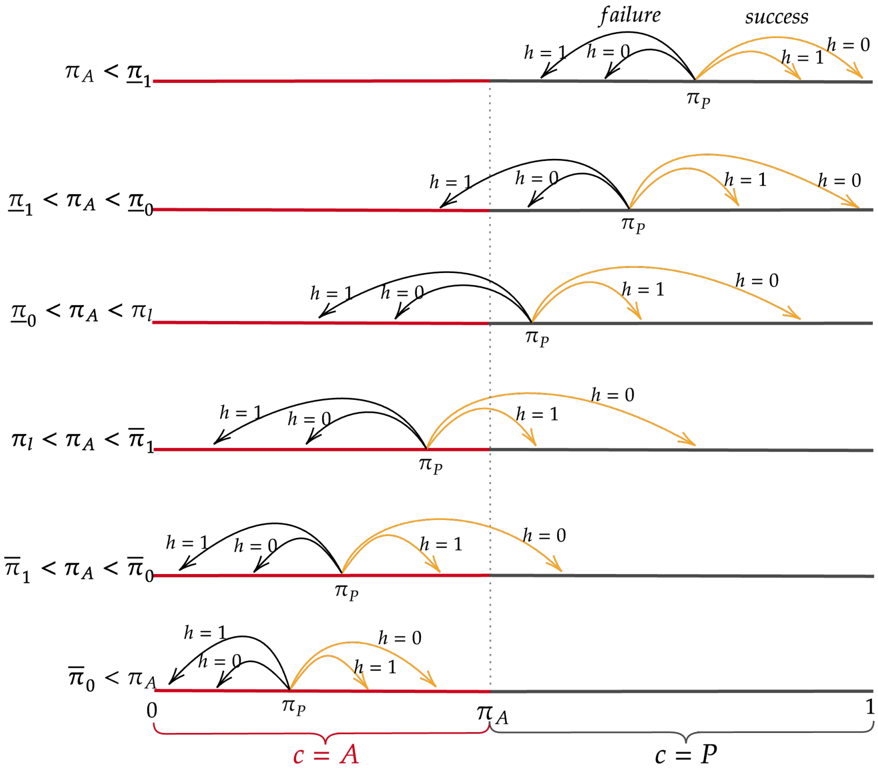
failure
success
$\pi_A < \underline{\pi}_1$
$h = 1$
$h = 0$
$h = 1$
$h = 0$
π_P
$\underline{\pi}_1 < \pi_A < \underline{\pi}_0$
$\underline{\pi}_0 < \pi_A < \pi_l$
$\pi_l < \pi_A < \overline{\pi}_1$
$\overline{\pi}_1 < \pi_A < \overline{\pi}_0$
$\overline{\pi}_0 < \pi_A$
0
π_A
1
$c = A$
$c = P$

Figure 2: Committee's selection given π_P and π_A.

failed execution more likely. Since by Assumption 1 the agent values the policy less than the office rent, he sabotages. Hence, we see how selection concerns can produce disunity or tension in an organization, an insight long present in the literature (e.g., Egorov and Sonin, 2011; Zhou, 2023; Geelen and Hajda, 2024).

However, case (*iii.*) illustrates how selection concerns may also incentivize the agent to help. In (*iii.*)(a), the committee selects the principal unless the policy fails and sufficient blame is attributed to her. That is, she must fail 'big'. So, the agent's selection concerns incentivize him to help and shift the blame for failure to the principal. In case (*iii.*)(b), the committee only selects the principal after success if sufficient credit is attributed to her. That is, she must succeed 'alone'. So, the agent's selection concerns incentivize him to help and share credit for success.[11] The next proposition formalizes the above discussion.[12]

Proposition 1 *A unique equilibrium exists almost everywhere.*

(i.) If $\pi_A \leq \underline{\pi}_1$*,* $h^* = 1$ *if and only if* $\varphi \geq 0$*. Otherwise,* $h^* = 0$*. The committee always selects the principal.*

(ii.) If $\underline{\pi}_1 < \pi_A \leq \underline{\pi}_0$*,* $h^* = 1$ *if and only if*

$$\varphi \geq \underline{\varphi}_1 \equiv -\frac{1 - \boldsymbol{f}(\pi_P, 1)}{\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0)} b.$$

Otherwise, $h^* = 0$*. If* $h^* = 1$*, the committee selects the principal if and only if the policy succeeds. If* $h^* = 0$*, the committee always selects the principal.*

(iii.) If $\underline{\pi}_0 < \pi_A \leq \overline{\pi}_1$*,* $h^* = 0$ *and the committee selects the principal if and only if the policy succeeds.*

(iv) If $\overline{\pi}_1 < \pi_A \leq \overline{\pi}_0$*,* $h^* = 1$ *if and only if*

$$\varphi \geq \underline{\varphi}_2 \equiv -\frac{\boldsymbol{f}(\pi_P, 0)}{\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0)} b.$$

Otherwise, $h^* = 0$*. If* $h^* = 1$*, the committee never selects the principal. If* $h^* = 0$*, the committee selects the principal if and only if the policy succeeds.*

(v) If $\overline{\pi}_0 < \pi_P$*,* $h^* = 1$ *if and only if* $\varphi \geq 0$*. Otherwise,* $h^* = 0$*. The committee never selects the principal.*

This result offers a new insight into the strategic incentives created by rank-order competition. A common result in this literature is that increasing the prize spread increases the incentive to sabotage (Lazear, 1989; Dye, 1984; Falk et al., 2008). However, the following corollary shows the reverse.

[11] In the Supplementary Appendix, we analyze a version of the model where the committee does not observe the agent's decision. We show that if his decision is not publicly observable, he no longer helps to share credit for success or avoid blame for failure.

[12] We denote an equilibrium action by a raised $*$.

Corollary 1 *The thresholds $\underline{\varphi}_1$ and $\underline{\varphi}_2$ are strictly decreasing in b.*

The key to this difference is that, in our model, sabotage affects the probability of successful execution but also the *information* revealed by the outcome. Sometimes, the agent helps due to succession concerns. When this is the case, his incentive to help increases as b increases.

5 Misaligned Policy Preferences ($\bar{x} > 0$)

As illustrated by the discussion of the mayor and governor in the introduction, the principal and agent may have different policy preferences. In this section, we explore this possibility. Specifically, we assume that while the principal and agent have some common interest in the successful execution of the policy, they also disagree about some aspects.

Assumption 2 $\bar{x} > 0$ *and* $\varphi > 0$.

In addition, to focus on the most interesting region of the parameter space, we assume that the extent of disagreement between the principal and agent is sufficient for the principal to face a trade-off between obtaining the agent's help and choosing her most preferred agenda. Formally, this means:

Assumption 3

$$\bar{x} > \varphi + \frac{\max\{\boldsymbol{f}(\pi_P, 0), 1 - \boldsymbol{f}(\pi_P, 1)\}}{\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0)} b.$$

5.1 Analysis

Our central insight, Proposition 2, relates to the principal's incentive to obtain the agent's help when doing so increases the probability she is replaced. Before discussing Proposition 2, we describe equilibrium behavior in a series of *Results*. These *Results* are corollaries of Proposition A in the Appendix, which characterizes the unique equilibrium.

We begin by considering the principal's agenda-setting decision and how it affects the agent's decision to help. The principal can always choose her preferred policy, i.e., $x = \bar{x}$, but by Assumption 3, this will come at the expense of sabotage. Naturally, the principal's alternative option is to make a policy concession—by choosing a smaller x—to obtain the agent's help.

No Policy Concessions due to Selection Concerns. Consider the region of the parameter space where the committee selects the agent if and only if the policy fails, even if failure is partly due to sabotage. (case *(ii.)* in Lemma 1). Then, there is no policy concession the principal can make to obtain the agent's help since, by Assumption 1, the agent's office rent is greater than his policy payoff for any x. As a result, the principal chooses her preferred policy, $x^* = \bar{x}$.[13]

Result 1 *Consider $\underline{\pi}_0 < \pi_A \leq \overline{\pi}_1$, the committee selects the principal if and only if the policy succeeds. Then $x^* = \bar{x}$ and $h^* = 0$.*

Outside the region where $\underline{\pi}_0 < \pi_A \leq \overline{\pi}_1$, the principal can obtain the agent's help by making a policy concession.

Policy Concessions without Selection Concerns. If selection is independent of the outcome (case $(i.)$ in Lemma 1), the agent's decision only depends on the policy. He helps if and only if:

$$\mathbb{E}[u_A(x) \mid h = 1] = \boldsymbol{f}(\pi_P, 1)(\varphi - x) \geq \boldsymbol{f}(\pi_P, 0)(\varphi - x) = \mathbb{E}[u_A(x) \mid h = 0]. \tag{4}$$

Since the principal prefers a larger x, her optimal policy concession makes this inequality bind. Thus, in equilibrium, she chooses between $x^* = \varphi$, in which case the agent helps, and $x^* = \bar{x}$, in which case the agent sabotages.

Result 2 *Consider either $\pi_A \leq \underline{\pi}_1$, where the committee always selects the principal, or $\overline{\pi}_0 < \pi_A$, where the committee never selects the principal. If*

$$\mathbb{E}[u_P(x = \varphi) \mid h = 1] = \boldsymbol{f}(\pi_P, 1)2\varphi + b \geq \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x}) + b = \mathbb{E}[u_P(x = \bar{x}) \mid h = 0], \tag{5}$$

$x^* = \varphi$ and $h^* = 1$. Otherwise, $x^* = \bar{x}$ and $h^* = 0$.

Policy Concessions Leveraging Selection Concerns. When the committee's selection depends on the outcome and the agent's decision (case $(iii.)$ in Lemma 1), selection concerns incentivize the agent to help. In particular, if the principal is selected unless she fails 'big'—i.e., with the agent's help—the agent is incentivized to help so that she shifts the blame for failure to the principal. Hence, he helps if and only if:

$$\begin{aligned}\mathbb{E}[u_A(x) \mid h = 1] = \boldsymbol{f}(\pi_P, 1)(\varphi - x) + (1 - \boldsymbol{f}(\pi_P, 1))b& \\ \geq \boldsymbol{f}(\pi_P, 0)(\varphi - x) = \mathbb{E}[u_A(x) \mid h = 0].& \quad (6)\end{aligned}$$

As a result, the principal's optimal concession is:

$$\hat{x}_1 \equiv \varphi + \frac{(1 - \boldsymbol{f}(\pi_P, 1))}{\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0)} b. \tag{7}$$

Similarly, if the committee only chooses the principal if she succeeds 'alone'—despite the agent's sabotage—the agent is incentivized to help so that she shares credit for success. By a

[13] This observation about the interaction between the principal's agenda-setting power and the agent's help is relevant to the literature on intra-party competition if we interpret the principal as the leader of a party and the agent as her heir apparent. If anything, existing research points to how the opportunity to sabotage the leader weakens her agenda-setting power (Izzo, 2024). In contrast, we show that a leader who is sabotaged in equilibrium chooses her preferred policy.

similar argument to the one above, we obtain the optimal concession:

$$\hat{x}_2 \equiv \varphi + \frac{\boldsymbol{f}(\pi_P, 0)}{\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0)} b. \tag{8}$$

It is clear that $\hat{x}_1$ and $\hat{x}_2$ imply a smaller concession than the one that makes (4) bind, i.e., $x = \varphi$. So the principal can make a smaller concession and obtain help because she *leverages the agent's selection concerns.*

Result 3 *(i.) Consider $\underline{\pi}_1 < \pi_A \le \underline{\pi}_0$, where the committee selects the principal unless the policy fails despite the agent's help. If*

$$\begin{aligned}\mathbb{E}[u_P(x = \hat{x}_1) \mid h = 1] &= \boldsymbol{f}(\pi_P, 1)(\varphi + \hat{x}_1 + b) \\ &\ge \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x}) + b = \mathbb{E}[u_P(p = \bar{x}) \mid h = 0],\end{aligned} \tag{9}$$

$x^* = \hat{x}_1$ and $h^* = 1$. Otherwise, $x^* = \bar{x}$ and $h^* = 0$.

(ii.) Consider $\overline{\pi}_1 < \pi_A \le \overline{\pi}_0$, where committee only selects the principal if the policy succeeds despite the agent's sabotage. If

$$\begin{aligned}\mathbb{E}[u_P(x = \hat{x}_2) \mid h = 1] &= \boldsymbol{f}(\pi_P, 1)(\varphi + \hat{x}_2) \\ &\ge \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x} + b) = \mathbb{E}[u_P(p = \bar{x}) \mid h = 0],\end{aligned} \tag{10}$$

$x^* = \hat{x}_2$ and $h^* = 1$. Otherwise, $x^* = \bar{x}$ and $h^* = 0$.

As condition (9) and (10) illustrate, a principal who leverages her agent's selection concerns faces a nuanced trade-off when inducing the agent to help: she obtains help with a smaller policy concession, but help comes with the cost of a decreased probability of selection. For a principal who the committee selects unless she fails 'big', obtaining help opens the door for the agent to be selected. For a principal who the committee does not select unless she succeeds 'alone', obtaining help ends any hope of being selected.

The Advantage of Leveraging Selection Concerns. One might conjecture that the principal is less inclined to obtain the agent's help when it comes at the cost of a decreased probability of being selected. However, the opposite is true: the principal has the greatest incentive to obtain help when she can leverage selection concerns. The next proposition formalizes this result.

Proposition 2 *For a given principal with prior π_P, consider an agent with prior π_A. If the principal weakly prefers to obtain help when π_A is such that she cannot leverage selection concerns—i.e., $\pi_A \le \underline{\pi}_1$ or $\overline{\pi}_0 < \pi_A$—she strictly prefers to obtain help when she can leverage selection concerns—i.e., $\underline{\pi}_1 < \pi_A \le \underline{\pi}_0$ or $\overline{\pi}_1 < \pi_A \le \overline{\pi}_0$.*

This result stems from a key difference between the agent's incentive to help due to a policy concession and his incentive to help due to selection concerns. Since the policy succeeds

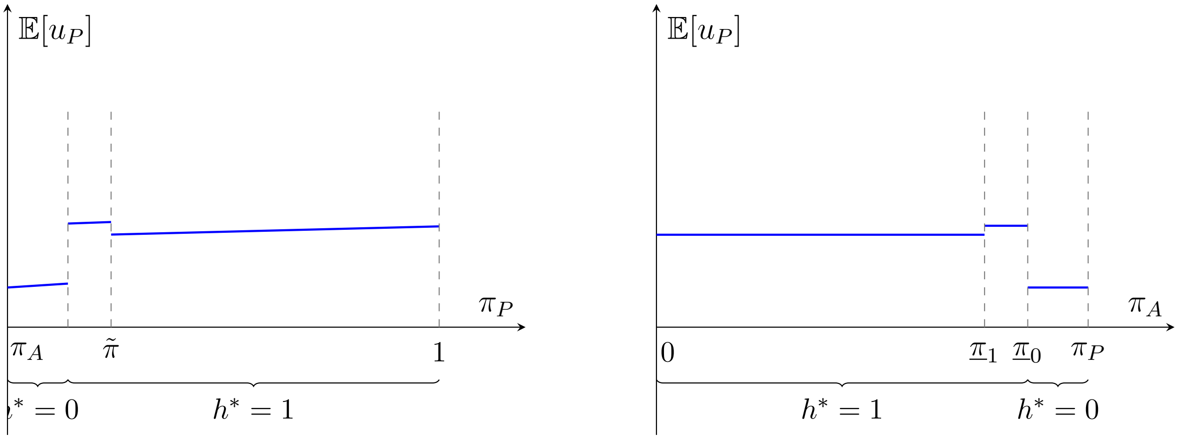

Figure 3: Left panel: Principal's expected utility as a function of $\pi_P \in [\pi_A, 1]$. Right panel: Principal's expected utility as a function of $\pi_A \in [0, \pi_P]$. We assume $f(\theta_P, h) = v + mh + r\theta_P$, where $v = \frac{1}{20}$, $m = \frac{1}{3}$, $r = \frac{6}{25}$, $\pi_P = \frac{1}{2}$, $\bar{x} = \frac{8}{5}$, $\varphi = \frac{1}{2}$, and $b = \frac{1}{2}$.

with positive probability despite sabotage, i.e., $\boldsymbol{f}(\pi_P, 0) > 0$, the agent benefits from the principal's policy concession regardless of whether he helps or sabotages. Things are different with the agent's incentive to help due to selection concerns. In this case, the agent only benefits if he helps and thus influences the updating of beliefs about the principal's quality. In other words, selection concerns create a contingent incentive for the agent, making it less costly for the principal to secure his help.

Building on this insight, we can now examine the principal's expected utility as a function of the agent she is paired with. First, whenever the principal can obtain the agent's help through some policy concession, she is better off than when she cannot. Second, whenever the principal can leverage selection concerns to obtain help she is better off than when she needs to rely solely on policy concessions. Taken together, these observations imply that the principal's expected utility is maximized when she can leverage selection concerns. Moreover, as long as the extent of disagreement between principal and agent is not too large, her expected utility is non-monotonic in the agent's expected quality. The following proposition formalizes this discussion.[14]

Proposition 3 *Consider the expected utility of a principal with prior π_P with respect to the agent's prior $\pi_A \in [0, \pi_P)$. The principal's expected utility is maximized when she can leverage selection concerns, i.e., $\underline{\pi}_1 < \pi_A \leq \underline{\pi}_0$. Furthermore, there exists a unique threshold χ such that for any $\bar{x} < \chi$, her expected utility is strictly non-monotone with respect to π_A.*[15]

This proposition has an interesting implication when we interpret the principal as the

[14] If $\boldsymbol{f}(\pi_P, 0) < \frac{1}{2}$, the result in Proposition 3 is true for $\pi_A \in [0, 1]$. If $\boldsymbol{f}(\pi_P, 0) > \frac{1}{2}$, the principal's expected utility is maximized for $\pi_A \in [0, 1]$ when $\overline{\pi}_1 < \pi_A \leq \overline{\pi}_0$.

[15] The threshold χ is the value of $\bar{x}$ above which the principal chooses her most preferred policy for any $\pi_A \in [0, \pi_P)$.

leader of an organization who gets to choose her presumptive successor, the agent. As the right panel in Figure 3 shows, if a principal with prior π_P is allowed to select an agent from a continuum of candidates with priors $[0, \pi_P)$, she will select neither the best nor the worst candidate, but an agent who will replace her if she fails 'big'. This insight unveils a novel rationale for why leaders may choose mediocre collaborators, a topic addressed by the political economy literature on 'mediocracy' (Caselli and Morelli, 2004; Mattozzi and Merlo, 2015).

Lastly, the left panel in Figure 3 shows the principal's expected utility is also non-monotonic with respect to her own expected quality, π_P. Generally, increasing her expected quality has a positive effect: it mechanically increases the success probability because the principal is more likely to be a high type and, also indirectly by making the agent's sabotage less likely. However, the principal's expected quality decreases at the point—denoted $\tilde{\pi}$ in the figure—above which she cannot leverage selection concerns and must rely only on policy concessions.

6 Conclusion

In this article, we presented a model of rank-order competition. In our model, a principal designs a policy, and her agent chooses whether to help execute it or sabotage it. We show that rank-order competition creates an incentive for the agent to help the principal. By helping, he shares credit for success and avoids blame for failure. In turn, the principal can leverage the agent's ambition, using it to obtain his support for policies closer to her ideal point. Indeed, we show that the principal attains her highest expected utility when paired with a "mediocre" agent whose selection concerns she can exploit—even if said agent replaces her in the event of failure.

Our model lends itself to many possible extensions. A natural one would be to explore what happens if the principal and agent interact repeatedly over an infinite horizon. This would allow us to study two interesting questions. First, how do the principal's and agent's behavior change over time? Second, what happens when the principal has the ability to fire the agent? We leave a formal examination of these questions to future work.

References

Asprey, R. B. (2008): *The Rise of Napoleon Bonaparte*, Hachette UK.

Barbaro, M. (2007): "Embattled Chief Executive Resigns at Home Depot," *The New York Times*, https://www.nytimes.com/2007/01/03/business/03cnd-depot.html.

Blanes i Vidal, J. and M. Möller (2007): "When should leaders share information with their subordinates?" *Journal of Economics & Management Strategy*, 16, 251–283.

Caselli, F. and M. Morelli (2004): "Bad politicians," *Journal of Public Economics*, 88, 759–782.

Chen, K.-P. (2003): "Sabotage in Promotion Tournaments," *Journal of Law, Economics, and Organization*, 19, 119–140.

Dessein, W. and L. Garicano (2023): "La Cordata: Loyalty in Tournaments," *Working Paper*.

Dewan, T. and D. P. Myatt (2007): "Scandal, Protection, and Recovery in the Cabinet," *American Political Science Review*, 101, 63–77.

——— (2010): "The Declining Talent Pool of Government," *American Journal of Political Science*, 54, 267–286.

Drago, R. and G. T. Garvey (1998): "Incentives for Helping on the Job: Theory and Evidence," *Journal of labor Economics*, 16, 1–25.

Dwyer, P. (2007): *Napoleon: The Path to Power*, Yale University Press, New Haven, CT and London.

Dye, R. A. (1984): "The Trouble with Tournaments," *Economic Inquiry*, 22, 147.

Egorov, G. and K. Sonin (2011): "Dictators and Their Viziers: Endogenizing the Loyalty–Competence Trade-off," *Journal of the European Economic Association*, 9, 903–930.

Falk, A., E. Fehr, and D. Huffman (2008): "The Power and Limits of Tournament Incentives," *Working Paper*.

Geelen, T. and J. Hajda (2024): "Succession," *Working Paper*.

Gieczewski, G. and C. Li (2022): "Dynamic Policy Sabotage," *American Journal of Political Science*, 66, 617–629.

Grow, B. (2007): "Out at Home Depot," *NBC News*, https://www.nbcnews.com/id/wbna16469224.

Heo, K. and E. M. Wirsching (2024): "Bureaucratic Sabotage and Policy Inefficiency," *Working Paper*.

Hirsch, A. V. and J. P. Kastellec (2022): "A Theory of Policy Sabotage," *Journal of Theoretical Politics*, 34, 191–218.

Izzo, F. (2024): "With Friends like These, Who Needs Enemies?" *The Journal of Politics*, 86, 835–849.

Kang, M. (2022): "Let presidents fail: Congressional deference to presidents as gambling on failure," *Research & Politics*, 9, 20531680221093435.

Kavilanz, B. P. B. (2007): "Nardelli out at Home Depot," *CNN Money*, https://money.cnn.com/2007/01/03/news/companies/home_depot/.

Lazear, E. P. (1989): "Pay Equality and Industrial Politics," *Journal of political economy*, 97, 561–580.

Mattozzi, A. and A. Merlo (2008): "Political Careers or Career Politicians?" *Journal of Public Economics*, 92, 597–608.

——— (2015): "Mediocracy," *Journal of Public Economics*, 130, 32–44.

Prendergast, C. (1993): "A Theory of "Yes Men"," *The American Economic Review*, 757–770.

Roberts, A. (2014): *Napoleon: a life*, Penguin.

Zakharov, A. V. (2016): "The Loyalty-Competence Trade-off in Dictatorships and Outside Pptions for Subordinates," *The Journal of Politics*, 78, 457–466.

Zhou, C. (2023): "Last step to the throne: the conflict between rulers and their successors," *Political Science Research and Methods*, 11, 80–94.

A Appendix

Posterior Updating. To show condition (2), we obtain by Bayes rule the posterior beliefs:

$$\underline{\pi}_h = \frac{(1-f(1,h))\pi_P}{(1-f(1,h))\pi_P + (1-f(0,h))(1-\pi_P)} \text{ and } \overline{\pi}_h = \frac{f(1,h)\pi_P}{f(1,h)\pi_P + f(0,h)(1-\pi_P)}$$

By standard algebra,

$$\underline{\pi}_h = \frac{1}{1+\frac{1-\pi_P}{\pi_P}\frac{(1-f(0,h))}{(1-f(1,h))}} \text{ and } \overline{\pi}_h = \frac{1}{1+\frac{1-\pi_P}{\pi_P}\frac{f(0,h)}{f(1,h)}}$$

Hence, the condition (2) follows from the assumed information structure.

A.1 Proof of Lemma 1, Proposition 1, and Results 1, 2, and 3

Lemma 1 and Results 1, 2, and 3 follow directly from the following proposition. Proposition 1 follows by assuming $\bar{x} = 0$.

Let $\hat{x}_1$ and $\hat{x}_2$ be as defined in (7) and (8), respectively.

Proposition A *A unique equilibrium exists almost everywhere.*

(i.) If $\pi_A \leq \underline{\pi}_1$, when (5) holds, $x^* = \varphi$ and $h^* = 1$; otherwise, $x^* = \bar{x}$ and $h^* = 0$. The committee always selected the principal.

(ii.) If $\underline{\pi}_1 < \pi_A \leq \underline{\pi}_0$, when (9) holds, $x^* = \hat{x}_1$, $h^* = 1$, and the committee selects the principal if and only if the policy succeeds. Otherwise, $x^* = \bar{x}$, $h^* = 0$, and the committee always selects the principal.

(iii.) If $\underline{\pi}_0 < \pi_A \leq \overline{\pi}_1$, $x^* = \bar{x}$, $h^* = 0$, and the committee selects the principal if and only if the policy succeeds.

(iv) If $\overline{\pi}_1 < \pi_A \leq \overline{\pi}_0$, when (10) holds, $x^* = \hat{x}_2$, $h^* = 1$, and the committee never selects the principal. Otherwise, $x^* = \bar{x}$, $h^* = 0$, and the committee selects the principal if and only if the policy succeeds.

(v) If $\overline{\pi}_0 < \pi_A$, when (5) holds, $x^* = \varphi$ and $h^* = 1$; otherwise, $x^* = \bar{x}$ and $h^* = 0$. The committee never selects the principal.

Proof:

After updating, the committee selects the agent it believes to be more likely the high type. In case of indifference, we assume the committee retains the leader.

If $\pi_A \leq \underline{\pi}_1$ or $\overline{\pi}_0 < \pi_A$, the agent helps if and only if (4) is satisfied; i.e., $\varphi \geq x$. Since $x = \varphi$ maximizes the principal's utility conditional on the agent helping, $x^* \in \{\bar{x}, \varphi\}$. The principal chooses $x^* = \varphi$ if and only if (5) is satisfied.

If $\underline{\pi}_1 < \pi_A \leq \underline{\pi}_0$, the agent helps if and only if (6) is satisfied. Since $x = \hat{x}_1$ maximizes the principal's utility conditional on the agent helping, $x^* \in \{\bar{x}, \hat{x}_1\}$. The principal chooses $x^* = \hat{x}_1$ if and only if (9) is satisfied.

If $\overline{\pi}_1 < \pi_A \leq \overline{\pi}_0$, the agent helps if and only if

$$\boldsymbol{f}(\pi_P, 1)(\varphi - x) + b \geq \boldsymbol{f}(\pi_P, 1)(\varphi - x) + (1 - \boldsymbol{f}(\pi_P, 0))b \;\Leftrightarrow\; \hat{x}_2 \geq x.$$

Since $x = \hat{x}_2$ maximizes the principal's utility conditional on the agent helping, $x^* \in \{\bar{x}, \hat{x}_2\}$. The principal chooses $x^* = \hat{x}_2$ if and only if (10) is satisfied.

If $\underline{\pi}_0 < \pi_A \leq \overline{\pi}_1$, the agent helps if and only if

$$\boldsymbol{f}(\pi_P, 1)(\varphi - x) + (1 - \boldsymbol{f}(\pi_P, 1))b \geq \boldsymbol{f}(\pi_P, 0)(\varphi - x) + (1 - \boldsymbol{f}(\pi_P, 0))b,$$

which holds for no x, hence $x^* = \bar{x}$. □

A.2 Proof of Proposition 2

Suppose the principal weakly prefers to obtain help if $\pi_A \leq \underline{\pi}_1$ and $\overline{\pi}_0 < \pi_A$, i.e., $\boldsymbol{f}(\pi_P, 1)2\varphi \geq \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x})$. Then, if $\underline{\pi}_1 < \pi_A \leq \underline{\pi}_0$, the principal's expected utility from obtaining help is bounded below:

$$\begin{aligned}
&\boldsymbol{f}(\pi_P, 1)2\varphi + \boldsymbol{f}(\pi_P, 1)\left(\frac{1 - \boldsymbol{f}(\pi_P, 1)}{\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0)}b + b\right) \\
&\qquad\qquad \geq \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x}) + \boldsymbol{f}(\pi_P, 1)\left(\frac{1 - \boldsymbol{f}(\pi_P, 1)}{\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0)}b + b\right).
\end{aligned}$$

This implies the principal strictly prefers to obtain help when $\underline{\pi}_1 < \pi_A \leq \underline{\pi}_0$ since $\boldsymbol{f}(\pi_P, 0) > 0$ implies

$$\boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x}) + \boldsymbol{f}(\pi_P, 1)\left(\frac{1 - \boldsymbol{f}(\pi_P, 1)}{\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0)}b + b\right) > \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x}) + b.$$

A similar argument shows that if $\boldsymbol{f}(\pi_P, 1)2\varphi \geq \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x})$, the principal strictly prefers to obtain help when $\overline{\pi}_1 < \pi_A \leq \overline{\pi}_0$. □

A.3 Proof of Proposition 3

From Proposition 2 it follows that

$$\begin{aligned}
&\max\{\boldsymbol{f}(\pi_P, 1)(\varphi + \hat{x}_1 + b), \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x}) + b\} \\
&\qquad\qquad \geq \max\{\boldsymbol{f}(\pi_P, 1)(2\varphi + b), \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x}) + b\} \\
&\qquad\qquad\qquad\qquad > \boldsymbol{f}(\pi_P, 0)(\varphi + \bar{x} + b).
\end{aligned}$$

Hence, the principal's expected utility is weakly maximized when $\underline{\pi}_1 < \pi_A \leq \underline{\pi}_0$. Moreover, if (9) is satisfied, the principal's expected utility is non-monotone with respect to π_A. Condition (9) is satisfied as long as

$$\bar{x} < \frac{(\boldsymbol{f}(\pi_P, 1) - \boldsymbol{f}(\pi_P, 0))2\varphi}{\boldsymbol{f}(\pi_P, 0)} + \hat{x}_1 \equiv \chi.$$

Supplementary Appendix of "A Model of Selection and Policy Concerns"

A.1 Interpretation of Assumption (ii.)

Lemma I *Assumption (ii.) is satisfied if and only if* $\frac{\partial^2 f(\theta_P,h)}{\partial\theta_P\partial h}$ *is not too large.*

Proof:

Rearranging the conditions in $(ii.)$ reveals that the assumptions are equivalent to assuming

$$\frac{\partial^2 \ln(f(\theta_P,h))}{\partial\theta_P\partial h} \leq 0.$$

This condition is equivalent to

$$\frac{1}{f(\theta_P,h)}\frac{\partial^2 f(\theta_P,h)}{\partial\theta_P\partial h} - \left[\frac{\partial f(\theta_P,h)}{\partial\theta_P}\right]\left[\frac{\partial f(\theta_P,h)}{\partial h}\right]\frac{1}{(f(\theta_P,h))^2} \leq 0. \tag{11}$$

Since $f(\theta_P,h) \in [0,1]$ and is increasing in θ_P and h, (11) is satisfied as long as $\frac{\partial^2 f(\theta_P,h)}{\partial\theta_P\partial h}$ is not too large. □

A.2 Example Success Function

Lemma II *If* $f(\theta_P,h) = v + mh + (r_0 + r_1 h)\theta_P$, $v, r_1 \geq 0$, $r_0, m > 0$, $r_1 \leq \frac{mr_0}{v}$, *and* $v + m + r_0 + r_1 \leq 1$, $f(\theta_P,h) \in [0,1]$ *and satisfies (i.) and (ii.).*

Proof:

Note that $v \geq 0$ and $v + m + r_0 + r_1 \leq 1$ imply $f(\theta_P,h) \in [0,1]$. If $m, r_1 > 0$ and $r_0 \geq 0$, $f(\theta_P,h)$ is clearly strictly increasing in θ_P and h. Finally, since

$$\frac{\partial}{\partial h}\frac{f(1,h)}{f(0,h)} = \frac{r_1 v - r_0 m}{(me+v)^2},$$

and

$$\frac{\partial}{\partial h}\frac{1-f(0,h)}{1-f(1,h)} = \frac{r_0 m + r_1(1-v)}{(r_0 + h(r_1+m) + v - 1)^2},$$

$r_0 \geq \frac{r_1 v}{m}$ implies $\frac{f(1,h)}{f(0,h)}$ is weakly decreasing in h and $\frac{1-f(0,h)}{1-f(1,h)}$ is weakly increasing in h. □

A.3 Robustness: Agent's Type Affects Probability of Success

Suppose $f(\theta_P,h,\theta_A) = mh + r_P\theta_P + r_A\theta_A$, where $m + r_P + r_A \leq 1$, $m, r_A, r_P \geq 0$. In this section, we show that there exists a range of values of r_A such that:

(a) If $\pi_P = \frac{1}{2}$ and $\pi_A \leq \frac{1}{2}$, the agent is selected if and only if the policy fails despite the agent's help.

(b) if $\pi_P = \frac{1}{2}$ and $\pi_A \geq \frac{1}{2}$, the agent is selected unless the policy succeeds despite the agent's sabotage.

It follows from Proposition 1 that if one of these conditions is satisfied, the agent's selection concerns incentivize him to help. In the submission, we provide a Mathematica code that plots the region where such conditions are satisfied with respect to the agent's prior π_A and the agent's type effect parameter, r_A.

Proposition I *(i) If $\pi_A \leq \pi_P = \frac{1}{2}$ and condition*

$$r_A \in \left(\frac{r_P(1-\pi_A) + 2\pi_A - 1}{\pi_A}, \min \left\{ \frac{m(1-2\pi_A) + r_P(1-\pi_A) + 2\pi_A - 1}{\pi_A}, 1 - r_P - m \right\} \right)$$

is satisfied, the agent is only selected if the policy fails despite his help.

(ii) If $\pi_P = \frac{1}{2} \leq \pi_A$ and condition

$$r_A \in \left(\frac{m(1-2\pi_A) + r_P(1-\pi_A)}{\pi_A}, \min \left\{ \frac{m(1-2\pi_A) + r_P(1-\pi_A) + 2\pi_A - 1}{\pi_A}, 1 - m - r_P \right\} \right)$$

is satisfied, the agent is selected unless the policy succeeds despite his sabotage.

Proof:

Assume $\pi_P = \frac{1}{2}$. The agent is only selected if the policy fails despite her help if:

$$\Pr(\theta_P = 1|\overline{g}, 1) > \Pr(\theta_A = 1|\overline{g}, 1) \Leftrightarrow \frac{1-\pi_A}{\pi_A} \frac{m + \frac{r_P}{2}}{m + \frac{r_P}{2} + r_A} > \frac{m + \pi_A r_A}{m + \pi_A r_A + r_P}$$

$$\Pr(\theta_P = 1|\overline{g}, 0) > \Pr(\theta_A = 1|\overline{g}, 0) \Leftrightarrow \frac{1-\pi_A}{\pi_A} \frac{\frac{r_P}{2}}{\frac{r_P}{2} + r_A} > \frac{\pi_A r_A}{\pi_A r_A + r}$$

$$\Pr(\theta_P = 1|\underline{g}, 1) < \Pr(\theta_A = 1|\underline{g}, 1) \Leftrightarrow \frac{1-\pi_A}{\pi_A} \frac{1 - m - \frac{r_P}{2}}{1 - m - \frac{r_P}{2} - r_A} < \frac{1 - m - \pi_A r_A}{1 - m - \pi_A r_A - r}$$

$$\Pr(\theta_P = 1|\underline{g}, 0) > \Pr(\theta_A = 1|\underline{g}, 0) \Leftrightarrow \frac{1-\pi_A}{\pi_A} \frac{1 - \frac{r_P}{2}}{1 - \frac{r_P}{2} - r_A} > \frac{1 - \pi_A r_A}{1 - \pi_A r_A - r}.$$

These inequalities are satisfied by r_A satisfying $m + r_P + r_A \leq 1$ if

$$r_A \in \left(\frac{r_P(1-\pi_A) + 2\pi_A - 1}{\pi_A}, \right.$$
$$\min \left\{ \frac{m(1-2\pi_A) + r_P(1-\pi_A) + 2\pi_A - 1}{\pi_A}, \frac{m(1-2\pi_A) + r_P(1-\pi_A)}{\pi_A}, \right.$$
$$\left. \left. \frac{r_P(1-\pi_A)}{\pi_A}, 1 - r_P - m, \right\} \right)$$

If $\pi_A > \frac{1}{2}$,

$$\frac{m(1-2\pi_A) + r_P(1-\pi_A) + 2\pi_A - 1}{\pi_A} < \frac{r_P(1-\pi_A) + 2\pi_A - 1}{\pi_A},$$

and hence the set of r_A is empty. Consider then $\pi_A \leq \frac{1}{2}$. This implies the condition necessary for the agent to only be selected if the policy fails despite her help reduces to:

$$r_A \in \left(\frac{r_P(1-\pi_A)+2\pi_A-1}{\pi_A}, \min\left\{ \frac{m(1-2\pi_A)+r_P(1-\pi_A)+2\pi_A-1}{\pi_A}, 1-r_P-m \right\} \right), \tag{12}$$

which proves our first result.

The agent is selected unless the policy succeeds despite her sabotage if:

$$\begin{aligned}
\Pr(\theta_P=1|\overline{g},1) < \Pr(\theta_A=1|\overline{g},1) &\Leftrightarrow \frac{1-\pi_A}{\pi_A}\frac{m+\frac{r_P}{2}}{m+\frac{r_P}{2}+r_A} < \frac{m+\pi_A r_A}{m+\pi_A r_A + r_P} \\
\Pr(\theta_P=1|\overline{g},0) > \Pr(\theta_A=1|\overline{g},0) &\Leftrightarrow \frac{1-\pi_A}{\pi_A}\frac{\frac{r_P}{2}}{\frac{r_P}{2}+r_A} > \frac{\pi_A r_A}{\pi_A r_A + r} \\
\Pr(\theta_P=1|\underline{g},1) < \Pr(\theta_A=1|\underline{g},1) &\Leftrightarrow \frac{1-\pi_A}{\pi_A}\frac{1-m-\frac{r_P}{2}}{1-m-\frac{r_P}{2}-r_A} < \frac{1-m-\pi_A r_A}{1-m-\pi_A r_A - r} \\
\Pr(\theta_P=1|\underline{g},0) < \Pr(\theta_A=1|\underline{g},0) &\Leftrightarrow \frac{1-\pi_A}{\pi_A}\frac{1-\frac{r_P}{2}}{1-\frac{r_P}{2}-r_A} < \frac{1-\pi_A r_A}{1-\pi_A r_A - r}.
\end{aligned}$$

These inequalities are satisfied by r_A satisfying $m+r_P+r_A \leq 1$ if

$$\begin{aligned}
r_A \in \Bigg(& \frac{m(1-2\pi_A)+r_P(1-\pi_A)}{\pi_A}, \\
& \min\left\{ \frac{m(1-2\pi_A)+r_P(1-\pi_A)+2\pi_A-1}{\pi_A}, \frac{r_P(1-\pi_A)}{\pi_A}, \right. \\
& \left. \frac{r_P(1-\pi_A)+2\pi_A-1}{\pi_A}, 1-m-r_P \right\} \Bigg)
\end{aligned}$$

If $\pi_A < \frac{1}{2}$,

$$\frac{m(1-2\pi_A)+r_P(1-\pi_A)}{\pi_A} > \frac{m(1-2\pi_A)+r_P(1-\pi_A)+2\pi_A-1}{\pi_A}$$

and hence the set of r_A is empty. So suppose $\pi_A \geq \frac{1}{2}$. This implies the condition necessary for the agent to be selected unless the policy succeeds despite her sabotage reduces to:

$$r_A \in \left(\frac{m(1-2\pi_A)+r_P(1-\pi_A)}{\pi_A}, \min\left\{ \frac{m(1-2\pi_A)+r_P(1-\pi_A)+2\pi_A-1}{\pi_A}, 1-m-r_P \right\} \right),$$

which proves our second result. □

A.4 Robustness: Private Action

Assume that $\bar{x}=0$, $\varphi>0$, and the agent's decision to help or sabotage is not publicly observed. In this section, we show that the agent is no longer incentivized to help as a way to shift blame.

Proposition II *Suppose that $\bar{x}=0$, $\varphi>0$. For any pair of π_P and π_A, except if it is such* *that $\underline{\pi}_1 < \pi_P \leq \underline{\pi}_0$ or $\overline{\pi}_1 < \pi_P \leq \overline{\pi}_0$, the unique equilibrium is behaviorally identical to the unique equilibrium of the game where the agent's action is observed. Consider now*

(i.) If $\underline{\pi}_1 < \pi_P \leq \underline{\pi}_0$ the unique equilibrium is in mixed strategies: $h^* = 1$ with probability $\sigma_A^* = \frac{\pi_P - \underline{\pi}_0}{\underline{\pi}_1 - \underline{\pi}_0}$, $c^* = P$ if the policy succeeds, and $c^* = P$ with probability $\sigma_C = \frac{b-\varphi}{b}$ if the policy fails.

(ii) If $\overline{\pi}_1 < \pi_P \leq \overline{\pi}_0$, there exists an equilibrium that is behaviorally identical to the unique equilibrium of the game where the agent's action is observed, i.e., $h^* = 1$ and $c^* = A$. In addition, two other equilibria exist:

(a) $h^* = 0$ and $c^* = P$ if and only if the policy succeeds;

(b) $h^* = 1$ with probability $\sigma_A = \frac{\pi_P - \overline{\pi}_0}{\overline{\pi}_1 - \overline{\pi}_0}$, $c^* = P$ with probability $\sigma_C = \frac{\overline{\varphi}}{b}$ if the policy succeeds, and $c^* = A$ if the policy fails.

Proof:

We start with the cases where the agent's strategic considerations are the same as in the model where his action is observed. First, if $\pi_P \leq \underline{\pi}_1$, the committee selects the principal regardless of the policy outcome. Similarly, if $\overline{\pi}_0 < \pi_P$, the committee selects the agent regardless of the policy outcome. Hence, the agent helps since $\varphi > 0$. Second, if $\underline{\pi}_0 < \pi_P \leq \overline{\pi}_1$, the committee selects the principal if and only if the policy succeeds. Hence, the agent sabotages since $\varphi < b$.

Consider now the case $\underline{\pi}_1 < \pi_P \leq \underline{\pi}_0$, where the committee selects the principal unless the policy fails despite the agent's help. First, we show that there is no equilibrium such that the agent plays a pure strategy equilibrium. Suppose there exists an equilibrium in which the agent helps. Then, the agent is selected if there is failure. But then the agent helps if and only if

$$\boldsymbol{f}(\pi_P, 1)\varphi + (1 - \boldsymbol{f}(\pi_P, 1))b > \boldsymbol{f}(\pi_P, 0)\varphi + (1 - \boldsymbol{f}(\pi_P, 0))b,$$

which contradicts $b > \varphi$. Similarly, suppose there exists an equilibrium in which the agent sabotages. Then, the agent is never selected. But then the agent sabotages if and only if

$$\boldsymbol{f}(\pi_P, 0)\varphi \geq \boldsymbol{f}(\pi_P, 1)\varphi \Leftrightarrow \varphi \leq 0,$$

which is also a contradiction.

Finally, suppose in equilibrium such that $h = 1$ with probability $\sigma_A \in (0, 1)$ and $c = P$ with probability $\sigma_C \in (0, 1)$ if the policy fails. The committee is indifferent between selecting the principal or the agent if the policy fails if

$$\sigma_A = \frac{\pi_P - \underline{\pi}_0}{\underline{\pi}_1 - \underline{\pi}_0}.$$

The agent is indifferent between help and sabotage if

$$\sigma_C = \frac{b - \varphi}{b},$$

which is between 0 and 1 for all $\varphi \in (0, b)$.

Consider now the case $\overline{\pi}_1 < \pi_P \leq \overline{\pi}_0$, where the committee selects the agent unless the policy succeeds despite the agent's sabotage. We start with the equilibrium that is behaviorally identical to the unique equilibrium of the game with observable actions. An equilibrium where the agent helps exists if and only if

$$\boldsymbol{f}(\pi_P, 1)\varphi + b > \boldsymbol{f}(\pi_P, 0)\varphi + b \Leftrightarrow \varphi \geq 0$$

Now, we show that two other equilibria exist. First, an equilibrium exists where the agent sabotages if and only if

$$\boldsymbol{f}(\pi_P, 0)\varphi + (1 - \boldsymbol{f}(\pi_P, 0))b \geq \boldsymbol{f}(\pi_P, 1)\varphi + (1 - \boldsymbol{f}(\pi_P, 1))b.$$

This is always satisfied since $\varphi < b$

Second, suppose an equilibrium such that $h = 1$ with probability $\sigma_A \in (0, 1)$ and $c = P$ with probability $\sigma_C \in (0, 1)$ if the policy fails. The committee is indifferent between selecting the principal and agent after the policy succeeds if

$$\sigma_A = \frac{\pi_P - \overline{\pi}_0}{\overline{\pi}_1 - \overline{\pi}_0}.$$

The agent is indifferent between help and sabotage if

$$\sigma_C = \frac{\varphi}{b},$$

which is between 0 and 1 since $\varphi < b$. □