Dynamic Delegation

Parimal Bag[*] Prabal Roy Chowdhury[†]

This version: December 15, 2022

Abstract

A principal's leadership delegation pivots around the tradeoff between the experts' *talent*, i.e., ability to observe state-relevant signals, and *bias*, i.e., likelihood of making corrupt decisions. In a two-period framework with two experts, one more talented but less trustworthy relative to the other, we analyze a class of equilibria where the deputy shuts down critical information early on – *'yesman'* behavior. Somewhat surprisingly, early on talent is prioritized whenever the less talented expert is very likely unbiased, bias gap is large and the talented expert's honesty-adjusted talent level is high. In this case, if instead the more honest expert were selected as leader then, when truly honest, she *virtue-signals*, ignoring valuable information in project implementation to signal her honesty, undoing the value of having an unbiased leader. Second, we show that sometimes it is better for the principal that a talented deputy fully suppresses her information rather than make fully revealing recommendations, highlighting a tradeoff between first-round efficiency and screening.

JEL Classification: C72; D04; D73; D78.

Key Words: Organizations; politics; project choice; dynamic delegation; authority; talent; honesty; trust; signal respecting; virtue signalling; cut-throat; yesman; backstabbing; early-efficiency; screening.

[*]Department of Economics, National University of Singapore, Faculty of Arts and Social Sciences, AS2 Level 6, 1 Arts Link, Singapore 117570; E-mail: ecsbpk@nus.edu.sg

[†]Economics and Planning Unit, Indian Statistical Institute, New Delhi, India; E-mail: prabalrc1@gmail.com

1 Introduction

In a 2014 survey on leadership traits conducted by the Pew Research Center, about 84% of respondents say that it is honesty that matters the most, ranking it ahead of intelligence, decisiveness, organizational ability, compassion, innovativeness and ambition (Fig. 8). On the related subject of trust in leaders, another report describes, "Elected officials are at the bottom of the list, with 37% expressing at least a fair amount of confidence; 48% say they have not too much confidence and 15% express no confidence in elected officials to act in the public's best interests." Similar concerns have been raised about business leaders as well.[1,2] In fact such sentiments are not US specific, and are held by the public in many countries.[3] Despite this concern however the modern economics literature on leadership and authority lacks a formalization of the role of honesty and trust in decision making.

Consider decisions in organizations and politics as to the kind of projects to undertake and which policies to adopt. Should business be expanded in one region and not in another, or a new firm be acquired, whether to strike a trade deal (strengthening ties with one group while antagonizing another), or implement a controversial redistributive policy in a volatile and changing demography with election approaching? All of these require careful deliberations. When such decisions are made by experts who observe domain-specific soft information, it is natural to delegate authority based on merit, with the literature often identifying merit with talent, i.e. the experts' skills/accuracy in observing signals relevant to the decisions. This is broadly true of the extensive literature on organizations, delegation and expert advice.[4]

In this paper, we expand the definition of merit to also include *honesty* of the decision maker. Suppose the agents can be *corrupt*, and if given the authority can implement inefficient decisions in exchange for kickbacks and influence peddling.[5,6] With this broadened notion of merit, the problem of

---

[1] "Half or more of Americans think these influential people act unethically at least some of the time, ... Additionally, 77% believe this about the leaders of technology companies.. ...." https://www.pewresearch.org/politics/2019/09/19/why-americans-dont-fully-trust-many-who-hold-positions-of-power-and-responsibility/.

[2] See Fig. 9. https://www.pewresearch.org/politics/2019/07/22/how-americans-see-problems-of-trust/.

[3] See the falling trust in government and business in 28 countries with more than 36,000 respondents in the Edelman Trust Barometer 2022 at https://www.edelman.com/trust/2022-trust-barometer; Figs. 10 and 11.

[4] While Moisson and Tirole (2020) is the first paper on organizations that, to our knowledge, uses the term *merit*, we interpret the expert advice literature's use of the term talent to mean merit as well. In other contexts, especially in redistributive politics, merit always mean talent or productivity (e.g., Arrow et al., 2000; Almas et al., 2020).

[5] The Bofors scandal, the alleged kickback of millions of dollars to Indian and Swedish politicians by arms manufacturer Bofors in exchange for contracts worth billions of dollars, happened in the 1980s and 1990s. A google search on influence peddling bring up many similar instances of abuse of power by top-level public officials (finance ministers, premiers, presidents, senators) in countries such as Argentina, Brazil, Canada, France, Indonesia, Mexico, South Korea, and the United States. Arguably political leaders might be influenced in making fiscal policies (e.g., whether to impose corporate tax on windfall profits, or issue public debt when financing a freeze in energy prices, as observed recently in the UK) if they have connections in investment banking, energy sector, etc.; they may also adopt policies to favor certain industries in expectation of securing well-paying positions in those industries after retirement. Not surprisingly the issue of politicians' honesty is widely debated (Fisman, 2001; O'Neill, 2002; The Economist, 2009; Birch and Allen, 2010; Issacharoff, 2010; Smith, 2016).

[6] In corporate organizations, CEOs and managers are the relevant experts whose decisions can make or break company fortunes and their susceptibility to outside influence is equally plausible. Corruption by corporate managers, i.e., abuse of managerial position, does not come up in the news as much simply because it is not of general public interest. The lack of corruption data also limits academic research on the topic. Two notable exceptions are Mironov (2013) and Burguet and Che (2004). Mironov (2013) uses Russian banking transaction data to study managerial diversion of firm profits by $45,429$ companies that used *spacemen* (i.e., fly-by-night firms) to provide fake goods and services. Whereas

delegation becomes much more intricate: Should an expert be appointed to a decision leader's position based on superior talent, even though she might be corrupt, or should one rely on the expert who is perceived to be more honest? We raise this basic question by formulating decisions as a succession of adoption choices over pairs of binary projects, with the experts advising or in charge of implementing the projects. The *principal* wants to delegate authority to a leader to induce correct decisions.

How to detect a leader's (lack of) honesty? When kickbacks cannot be decisively proven, one way to discipline abuse of power is to draw inferences from the leader's actions, possibly taking away leadership from her and giving it to the other expert, her deputy advisor. Unfortunately however, this inevitably generates two further problems, both motivated by the incentive to sway the principal's leadership switching decision. The first is that of 'political correctness' (Morris, 2001) (or virtue signalling, as we call it), that may prompt even a non-corrupt expert to sometimes implement a conformist decision that is *ex ante* wrong. Second, initially the deputy may provide advice in a strategic fashion. In particular, even a non-corrupt deputy may keep her own counsel, not making any recommendation even when in possession of valuable information. In many hierarchical organizations it is well known that people in power often do not like dissenting voices. In a recent book, historian and former Whitehouse Aide Troy (2020) details how top politicians can sometimes be intolerant towards dissenting views.[7] We will offer a different explanation for a deputy's reticence.

In this paper, we will analyze a rich yet tractable dynamic cheap talk game that can formalize the tradeoff between talent and honesty from the principal's perspective. Talent of an expert is her ability to evaluate the unknown state. An expert is either biased or unbiased (informally, honest), with a biased agent deriving an additional unobserved benefit simply by implementing a specific project irrespective of its success. In order to maximize his objective, the principal has two tools at his disposal – initial delegation, and conditional replacement of a possibly corrupt leader in the next period, trading efficiency of decision making at the present against future (in)efficiency.

We build a one principal—two experts (agents) model without any commitment. The experts – one appointed by the principal as the leader and the other as the deputy – together make a choice from two projects in each of two rounds. While both the leader and the deputy each independently observe a signal conditional on the state and their talent, the leader has the ultimate authority in choosing the project.[8] Once the first period decision is made, the implemented project's success/failure is

---

Burguet and Che (2004) discuss many cases of bribery in the 1970s in both public and private procurement contracts, with 450 U.S. companies making questionable payments ($400 million) to foreign concerns, and losing $45 billion worth of foreign contracts to foreign competitors through graft in 1994-1995.

[7]In her book review, Burgess (2020) writes,

> "*Johnson's intolerance of internal conflict extended beyond infighting to dissenting views. The results for Johnson's Vietnam policy were disastrous. In the tightly controlled "Tuesday Group" of six foreign policy advisers with whom Johnson conferred about Vietnam, even Cabinet secretaries were kept from voicing concerns based on domestic considerations. Johnson bullied his senior staff, which, in turn, bullied subordinate staff Dissidents, who Troy says, "were derided, ganged-up on, or even dismissed.""*

Prendergast (1993) observes that a suboordinate worker reporting to the manager his finding about a project's value, or another worker's performance, is likely to be an 'yesman', i.e., second guess and conform to the manager's opinion.

[8]In party politics the advice of a deputy, or its lack of, is critical to the party's fortunes. Despite its importance such advice may or may not be heeded by the leader. Failing to perform during one's tenure as the leader mostly leads to a natural transition to the second-in-command. And this possibility puts the leader in a difficult situation to extricate any valuable information from the deputy. But it is also well known that having potential conflict of interest may bring out the truth (Shin, 1998; Dewatripont and Tirole, 1999; Krishna and Morgan, 2001).

revealed. Then using the documented advice by the deputy the principal updates his beliefs about the experts' honesty/bias, using it to decide what is best for period 2, whether to replace the leader or retain her. The leader in period 1 would like to continue in period 2 as (i) she might be corrupt and thus benefits from implementing her favorite project, or (ii) when she is not corrupt she would like to ensure efficient project choice in the future. In order to pose the tradeoff between talent and honesty more sharply, we consider a scenario where agent 1 is less talented but *ex ante* more honest relative to agent 2. Let ν_i denote the probability that expert i is honest. Further, for the discussion to follow consider a more comprehensive metric, the ***worth*** of an agent i: ex-ante accuracy of i's period 1 project choice as leader over and above the common prior on the unknown state.[9]

We focus on three key questions: (1) Can the deputy being initially completely silent, what we call *yesman* behavior, even be sustained as an equilibrium phenomenon? (2) When should a leader be replaced? (3) Should a principal appoint an expert who is more talented but also more likely to be corrupt as the initial leader, or should the choice fall on a more honest expert who is less talented? We shall answer the first question in the affirmative. Turning to the second question, we shall argue that the period 1 leader be replaced if and only if following her project choice she is perceived to be relatively less honest. Finally, and most importantly, we find that our third query has a somewhat non-obvious answer. In particular, talent wins, i.e. agent 2 is made the period 1 leader, whenever ν_1 is 'large', ν_2 is 'small', and agent 2 has relatively higher worth. Otherwise however, honesty can win.

We start by analyzing the continuation game in period 2, establishing that it has a 'sincere' equilibrium where the deputy discloses her signal truthfully (Proposition 1). Based on such advice an honest leader implements the project that is 'informationally efficient', whereas an unbiased leader always implements her biased project. Given this equilibrium in the continuation game, the principal selects the agent who is perceived to be more honest as the period 2 leader, with this perception being driven by the first period project choice only, not by its success or failure (Proposition 2).

Turning to the period 1 game, characterizing the set of perfect Bayesian equilibria (PBE) of a cheap talk dynamic game, with three strategic players and asymmetric information, is quite complex. For tractability we therefore focus on a class of equilibria that we however believe is quite 'natural', one where agents use pure strategies, the deputy makes a non-informative report in period 1 (which of course is the case of interest), and the period 2 game necessarily involves a sincere equilibrium. Note that while, as is standard, the second period game has multiple equilibria including a babbling one, focusing on the sincere equilibrium allows us to highlight equilibrium inefficiencies that are driven essentially by leader–deputy–principal strategic interactions in period 1, and not in period 2.

We solve for the natural PBE in two steps. First we fix agent 1 (resp. agent 2) as the period 1 leader, and solve for the Yesman-I (resp. Yesman-II) equilibrium of the continuation game (Propositions 4 and 5). With the payoffs thus determined, we next solve for the principal's period 1 leadership decision. Irrespective of which agent is the period 1 leader, we find that there are two common forms of inefficiency as far as project choice is concerned. The first involves *virtue signalling*, the unbiased agent ignoring her signal and necessarily implementing the project which is not favored by her biased self. The objective is to establish her credentials as an unbiased agent, and thereby retain period 2

[9]Despite its richness the worth of an expert will be less obvious to recognize as it is endogenously determined in equilibrium interactions, even if it depends on the two fundamentals of talent and bias.

leadership. Note that such behavior is more attractive if the current deputy is likely to be biased, because virtue signalling prevents such a deputy from gaining period 2 leadership. The other is *cut-throat* implementation, the biased leader necessarily implementing her favored project irrespective of her signal. This is more attractive if the current deputy is likely to be honest, since, given that the period 2 equilibrium is sincere, the cost of losing leadership would be smaller.

We start by discussing equilibria where the regime-switching rule follows a *natural* pattern – replacing the current leader if and only if she implements the favored project of a biased leader.[10] One important intuition is that given this regime-switching rule, period 1 and period 2 efficiencies are negatively related. If, say, the period 1 equilibrium involves both biased and unbiased leaders being signal respecting, then that improves period 1 efficiency, but reduces the efficacy of screening since, depending on her signal, and hence project choice, an honest leader may be screened out, whereas a biased leader may retain leadership. On the other hand, suppose the honest leader is virtue signalling, whereas the biased leader is cut-throat. Then period 1 implementation is inefficient with project choices unresponsive to signals, but screening is quite efficient as the honest agent is never screened out, whereas the biased leader is. This tradeoff makes the choice of period 1 leadership a challenging task.

We next discuss our third key result, that talent wins (agent 2 is made the period 1 leader) whenever ν_1 is large, ν_2 is either small, or intermediate, and the *worth* of agent 2 exceeds that of agent 1. We shall argue that these results follow from the interaction of two factors, the possibility of virtue signalling, and the negative tradeoff between period 1 and period 2 surplus. These results (Proposition 6)[11] are somewhat counter-intuitive because it may be argued that agent 1, who is more likely to be honest, and therefore possibly more likely to invest in the 'right' project, has some claim to period 1 leadership in particular if the talent levels are not too far apart. But note that the surplus from appointing either agent as the leader depends not just on her two attributes, honesty and talent, but also on her equilibrium project choice, which in turn depends on the level of honesty of the *other* agent. In particular, whenever ν_2 is either small, or intermediate, the Yesman-I equilibrium involves the honest leader virtue signalling, so that the period 1 surplus is likely to be quite low in case honesty wins. The other key intuition comes from the aforementioned tradeoff between period 1 and period 2 surplus. Thus, while the period 1 payoff under the Yesman-II equilibrium is not too large (because unbiased leader will be signal respecting whereas biased leader will be cut-throat and the ν_2 is not large), any Yesman-I equilibrium that has a larger period 1 surplus will also tend to have a lower period 2 surplus. Finally, whenever the period 1 screening efficacy is the same irrespective of whether talent or honesty wins, period 1 project efficiency which depends on the worth of the two agents is the determining factor. Thus, for these parameter regions talent wins.

We next parse these intuitions in greater details. First suppose ν_1 is large, ν_2 is small, and agent 2 has a higher worth. As discussed earlier, under the Yesman-I equilibrium only the biased leader is signal respecting, whereas under the Yesman-II equilibrium only the honest leader is. Thus agent 2 being the leader leads to higher period 1 surplus whenever the worth of agent 2 exceeds that of

[10]We shall discuss other regime-switching rules later.

[11]The discussion of the intuitions at this stage will be partial, restricted to the natural regime-switching rule. Proposition 6 will include a possibility where, under Yesman-I (i.e., more honest expert is the leader), there will be no regime switch following the implementation of the biased agent 1's favorite project; see Proposition 4.2. The discussion of the intuitions for the complete Proposition 6 is left for Section 6.

agent 1. Further, the screening efficiency turns out to be the same irrespective of who is the leader; the intuitions are explained following Proposition 6. Hence aggregate surplus is higher as well in case talent wins.

Next suppose ν_1 is large, and ν_2 is intermediate. With ν_2 intermediate, the Yesman-I equilibrium still involves virtue signalling by an honest leader. Further, the biased leader is cut-throat, so that all signals are ignored. Hence the period 1 surplus is higher if talent wins. Of course, if honesty wins, then screening is more efficient since a biased period 1 leader is necessarily screened out. Thus there is a tradeoff between period 1 and period 2 surplus. However, because of two reasons, the period 1 surplus effect dominates. First, period 1 payoffs are not too small (given that under the Yesman-II equilibrium the honest leader is signal respecting), and second, in case talent wins the signal received by an honest agent 2 is not lost, which is an important driver of period 1 surplus given that ν_2 is intermediate. Hence aggregate surplus is higher if talent wins.

On the other hand, honesty will win when, under the Yesman-II equilibrium, all informative signals are lost (so that ν_1 is at an intermediate level), and the Yesman-I equilibrium involves only the biased leader being signal-respecting (so that ν_2 is small) (Proposition 7.1). Given the nature of the two equilibria, period 1 surplus is higher if honesty wins. Further, while screening is more efficient if talent wins (since a biased period 1 leader is necessarily screened out), the period 1 effect dominates because, with ν_1 intermediate, the fact that a biased leader is signal respecting in case talent wins, adds significantly to the period 1 surplus differential.

There are other surplus comparisons but we will focus only on the case where both ν_1 and ν_2 are large so that the honesty *differential* is small. Depending on parameter values we find that either honesty or talent can dominate (Proposition 8). The case of ν_1 and ν_2 both being small, or both being intermediate generate very similar results that we do not report. Another comparison that we leave out is when Yesman-I equilibrium will have the principal using a "perverse" regime-switching rule, replacing period 1 leader provided she implements the project different from the one favored by her biased self. This leads both types of leader to always choose the favored project. Such an equilibrium is unappealing. The results for these two omitted cases appear in a working paper version (Bag and Roy Chowdhury, 2022).

Finally, we consider a class of equilibria which is the polar opposite of Yesman equilibria: Informative Advice equilibria where the period 1 deputy necessarily reports her signal truthfully (Proposition 9). The knee-jerk response would be to say that, fixing parameter values and period 1 leadership choice, such equilibria would necessarily dominate the corresponding Yesman equilibrium in terms of expected aggregate surplus. To our surprise, we find that this is not always the case (Propositions 10). The result hinges on a tradeoff between period 1 and period 2 efficiency. Under the informative advice equilibrium project choice in period 1 is efficient, but screening for period 2 is worse relative to the Yesman-I equilibrium. This follows since *information revelation* by the talented deputy has a cost as the biased leader can hide behind the deputy's advice and has little incentive to maintain a reputation for honesty. Interpreted in a more colorful language, we thus uncover a beneficial role of *backstabbing*: both in politics and organizations this so-called nefarious behavior, withholding of valuable information by the deputy, may help to remove potentially corrupt leaders. This result adds nuance to the organizational economics literature that views sabotage, a form of backstabbing, only as an efficiency worsening activity (Konrad, 2000; Chen, 2003).

■ **Literature review.** The framework and the analysis in this paper advances the rich literature on decision making in organizations with cheap talk advice and delegation. The power of cheap talk was analyzed in a sender–receiver/principal–agent game by Crawford and Sobel (1982), while Aghion and Tirole (1997) and Dessein (2002) observed that the principal, rather than making the decision himself, may want to delegate the decision authority to the expert.[12]

Our principal faces the question not about whether to delegate or not, but to which expert to delegate.[13] And delegation is not just once-for-all, future delegation will depend on the interim choices following the initial delegation.[14] The subject of cheap talk advice and dynamic project implementations is sparse.[15] Even rarer, none to our knowledge, is any study of the multi-period interaction between experts' talents and private biases and managerial corruption, and how this interaction shapes beliefs for future decisions. This interaction to ensue down the line is an integral part of the principal's scheme of dynamic delegation. In this sense, we go beyond the existing literature's central focus on binary, either/or choice between delegation and cheap talk advice.

The paper that most closely relates to ours is Morris (2001). He pointed out that an informed advisor, whose interests could well be aligned with that of the policy maker, faces the problem of making a truthful recommendation for reasons of reputational concerns.[16] A private bias of a social scientist might color her recommendation, so if her true belief is that affirmative action is an ill-conceived policy to address racism and recommends anti-affirmative action policy, it may lend weight to the misperception that she might be biased (i.e., racist), and her recommendation in the future for other policy decisions that could be linked to the bias won't be taken seriously. So the advisor is incentivized to lie and recommend affirmative action, i.e., be politically correct. A similar force is also in play in our analysis, in influencing a leader's project choice decision, compromising short-term efficiency. But the same short-term inefficiency can be a blessing for future efficiency. With these countervailing forces in play, how should the authorities then choose a leader from experts of differing perceived biases and known talents is a natural and significant issue to study, the answer to which is not clear from Morris (2001) and the follow-on literature. We advance this important research agenda.

Our problem can also be seen as an extension of Aghion and Jackson (2016), who consider a principal whose objective is to learn about a leader's credentials. They view an agent's talent/competence to be the key characteristic, so their principal wants to retain a competent leader and fire the incompetent one. The focus therefore is on optimal replacement decisions when the appointed leader

[12]Decision making in organizations using cheap talk advice and career concerns have appeared in some of the early works including Ottaviani and Sorensen (2006), Levy (2007), and Kartik (2009), and more recently, among others, Agastya et al. (2014), Bag and Sharma (2019). The experts' payoffs in the current formulation are more direct, linked to project and corruption proceeds, than in career-concern models.

[13]Holmstrom (1984) is an early paper pointing out the value of delegation.

[14]In a principal-agent model, Baker, Gibbons and Murphy (1999) viewed delegation as a "loaned authority". In their uninformed principal variant, the authors observe that the subordinate's reputation for using the delegated authority appropriately will be an important consideration because the authority can be retracted. In our paper, the retracted authority will be passed onto a second expert.

[15]Sobel (1985), Benabou and Laroque (1992), Morris (2001) analyze cheap talk in multi-period models. Golosov et al. (2014) shows the possibility of full information revelation in a finite period dynamic version of Crawford-Sobel game.

[16]That perverse influence of reputation can be quite bad even when the concerned player is there in the long term providing some service, such as a medical practitioner dealing with patients, has been shown by Ely and Valimaki (2003).

has the choice of a safe but uninformative action and an informative, risky action. Different from Aghion and Jackson, in our model competence is not the only concern, as honesty/bias also matters and talent might even be of secondary importance for leadership. We also incorporate strategic interactions between the leader and the deputy; Aghion and Jackson abstract from these interactions by not allowing for any deputy, instead allowing the replacement to be drawn from outside.

Finally, there are a number of related papers in the broad domain of project implementation, by collective decision in teams or through delegation, and power dynamics. In two contributions, Alonso and Matouschek (2007; 2008) consider a principal optimally choosing delegation of decision making to a biased but better informed agent manager, from threshold delegation (e.g., investment not to exceed a specified level), menu delegation (restricting agent's choice to a finite number of projects) to centralization, depending on to what extent the principal can commit not to overrule the agent. Armstrong and Vickers (2010) studied what kind of projects the principal should allow the agent to choose from when the agent's choice set (of available projects) is not fully known to the principal. Bester and Krahmer (2008) study, in a static setting, the tension between the structure of decision making rights (or authority over project choice) and efforts of the team members. Blanes i Vidal and Moller (2016) consider the dual aspects of project selection and project execution in a team, suggesting in order to keep the morale of its team members high so they exert efforts (motivation) there could be insufficient (verifiable rather than cheap talk) communication during the project selection stage (adaptation) when, ex ante, one of two projects is considered to be of higher quality. Li et al. (2017) analyze the dynamic incentives in organizations among prospective decision makers – the lure of power/authority in the future motivates the incumbent decision maker to make good/honest use of power today (present).[17] The link between our experts' interests in future leadership and the strategic decisions in project choice and advice giving in the present is similar to Li et al.'s power dynamics and the related incentives. The difference is that in our case the leader in the initial phase doesn't necessarily get rewarded for choosing a project that is proven, ex post, to be the correct choice. Also, an *ex-ante* correct choice is often penalized if it happens to be the perceived bias type's favorite project.

After presenting the model in the next section, the core analysis is developed in Sections 3–7. The main proofs are included in Appendix A. Supplementary Appendix B contains the analysis of the informative advice equilibria. Additional results and proofs are reported in Bag and Roy Chowdhury (2022). Mathematica programs used in constructing numerical illustrations of the various equilibria presented in the Figures can be made available on request.

[17] In a laboratory experiment, Fehr, Herz and Wilkening (2012) found that in project choice decisions the disutility from being overruled by the subordinate agent prevented in many cases delegation that would have been economically beneficial to the principal. In our setting, benefits and losses are all pecuniary, with the principal solving a very different problem: not whether to delegate or not but how to navigate the dynamic delegation path.

2 Model

A team of two agents (or experts), a leader (ℓ) and a deputy (d), has to adopt two projects, one after another, from a selection of two distinct pairs of projects over two rounds (or periods): either a_1 or b_1 in round 1, and either a_2 or b_2 in round 2. The project outcome, success or failure, depends on an unknown state realized independently in each round.

The realizations of project outcomes are independent across periods. Within each round, while *ex ante* the two projects are identical, *ex post* one project will prove to be superior.

The team's owner, the principal (denoted P), decides who will lead the team in period 1, and if the leader should be changed in period 2. All parties – the principal and the agent – are risk neutral with regard to any payoffs that accrue from project implementation including its success/failure. Moreover, all agents have a discount factor of 1, valuing future payoffs at par with that in the current period.

In round 1 the agent in charge of the team solicits advice from her deputy about which of the two projects a_1 or b_1 should be implemented. The advice is not binding as the leader has the real authority over project implementation. If the implemented project is successful, it yields a gross value $V_1 > 0$ and failure yields zero.[18] Likewise in round 2, the leader (who need not be the leader from round 1) again asks for the deputy's advice and implements a_2 or b_2 that yields $V_2 > 0$ on success and zero following failure. Any value resulting from a success will be split according to a fixed sharing rule $s_P, s_\ell, s_d > 0$, $s_P + s_\ell + s_d = 1$, where s_i denotes the share of i, $i = P, \ell, d$. We further impose a simplifying restriction that $s_\ell = s_d = s$ (say). Exogenous identical shares in both periods for the leader and the deputy is the only contractual commitment our principal makes.[19]

In addition, an agent i may be *unbiased* (i.e., honest[20] or non-corruptible, denoted N) or *biased* (i.e., corruptible, denoted $\neg N$), a characteristic we call $\kappa_i \in \{N, \neg N\}$ immutable throughout, with $\Pr(\kappa_i = N) = \nu_i$ and $\Pr(\kappa_i = \neg N) = 1 - \nu_i$, $i = 1, 2$. If agent 1 is biased, she derives extra private gains, $\zeta > 0$, merely by implementing projects a_1 and a_2 as the leader; likewise, if agent 2 is biased she derives similar gains $\zeta > 0$ by implementing projects b_1 and b_2 as the leader. There is no such extra gain to an agent if she is not the leader or if she is unbiased. It could be that the leader can exchange favors with outside parties by implementing specific projects and from which the involved parties stand to gain.

Project m's $(= a_t, b_t)$ success probability, $\gamma(m)$, depends on the true state, ω_t $(= a_t, b_t)$, as follows:

$$\gamma(m) = \begin{cases} 1, & \text{if } \; m = \omega_t \\ 0, & \text{if } \; m \neq \omega_t \end{cases}, \text{ where}$$

$$q_t = \Pr(\omega_t = a_t) = 1/2, \quad q'_t = \Pr(\omega_t = b_t) = 1/2, \text{ for } t = 1, 2. \tag{1}$$

Note that the project's success or failure precisely reveals the true state, and so, ex ante, projects a_t and b_t are of identical quality.

[18]What matters for qualitative predictions is the difference in the values from success and failure.

[19]In political applications, value generated from a successful project can be collective party benefits.

[20]We use the terminology, honest, in a broad sense. A leader may be unbiased/honest, yet may implement projects non-truthfully which can be interpreted as dishonesty. To avoid this ambiguity, throughout we will use the terminology, unbiased, rather than honest when referring to the agent's bias type. But when making relative comparisons such as more honest or less honest, we will understand it to refer to the relative probability of an agent being biased.

At the start of both the periods, agent i may or may not obtain an informative signal σ_i about the state of the world, with the signal generating technology differing across the two experts, though it remains identical across the two periods. Further, this technology is conditionally independent across the two experts. In every period, agents observe their signal independently, each observing no signal, $\emptyset$, with probability $0 < \epsilon < 1$, or an informative signal, α or β, with probability $1 - \epsilon$, where α and β are informative though noisy signals about which project is more likely to succeed. The distribution of the informative signal is determined by the agent's talent (or ability), τ_i, in predicting the true state of the world in each period. Formally, let

$$\tau_i = \Pr(\sigma_i = \alpha|\omega = a) = \Pr(\sigma_i = \beta|\omega = b),$$

where $\tau_1 = \xi(1 - \epsilon)$, $\tau_2 = \lambda(1 - \epsilon)$.

We are interested in the tradeoff between honesty and talent, and focus on the case where one agent (agent 1) is more talented, but less likely to be honest. We therefore have the following assumption:

Assumption 1 (a) *Agent 1 is of lesser talent than agent 2:* $\lambda > \xi > \frac{1}{2}$.

(b) *Agent 1 is more honest, ex ante:* $\nu_1 > \nu_2$.

The principal (he) is only interested in maximizing his share of the aggregate expected project returns (surplus), and does not take either the leadership gains, or the bias benefits into accounts. He chooses one of the agents to be period 1 leader, then selects the period 2 leader based on period 1 project choice, deputy recommendation, and project outcome (success or failure) to maximize his continuation payoff. Thus the continuation rule is one of *non-commitment.*

The model parameters $\{q_t, V_1, V_2, \lambda, \xi, \nu_1, \nu_2, \zeta\}$ are common knowledge. We will allow V_1 and V_2 to differ. We maintain the following assumption for much of the paper:

Assumption 2 (a) $\zeta > \frac{\lambda+\xi-1}{\lambda\xi+(1-\lambda)(1-\xi)}(sV_2)$; (b) $\zeta > (2\lambda - 1)(sV_1)$.

As we shall later see, part (a) ensures that the bias (or corruption) benefits are sufficiently large in that in period 2 a biased leader always opts for her pet project. Part (b) of the assumption is to rule out the case where even the corrupt but talented leader would always implement informationally efficient project in period 1.

Throughout, we will consider the following communication protocol Ψ by the deputy. In each period the deputy can submit a recommendation about the project to be implemented ($\hat{\alpha}$ implying a_1 or a_2, and $\hat{\beta}$ implying b_1 or b_2), together with whether the recommendation is strong ($\hat{\lambda}$) or weak ($\hat{\xi}$), or submit a null report ($\emptyset$) meaning neither project recommendation nor communication about the level of talent. Formally,

$$R_{j,t} : \mathcal{T}_d \equiv \{(\alpha, \lambda), (\beta, \lambda), (\alpha, \xi), (\beta, \xi)\} \longrightarrow \{(\hat{\alpha}, \hat{\lambda}), (\hat{\beta}, \hat{\lambda}), (\hat{\alpha}, \hat{\xi}), (\hat{\beta}, \hat{\xi}), \emptyset\}, \tag{2}$$

where j is the index for the deputy (agent 1 or 2) and t for period 1 or 2. The null report means the deputy is entitled to keep her own counsel.

A sketch of the extensive form game of leader selection, reporting and decision making is summarized below, where decision points $\{T_{11}, ..., T_{14}\}$ make up period 1, $\{T_{21}, ..., T_{24}\}$ constitute period 2 and T_{inter} as the interim stage of updating by the principal and the agents of agents' bias type. More formal details such as set of types, strategies, beliefs, and payoffs will be provided as we analyze the game later on.

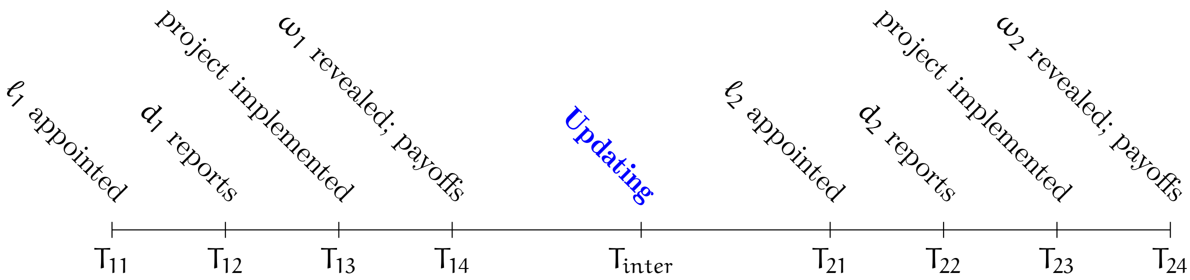


Figure 1: Time line – leadership choice without commitment

We solve the game by applying the standard solution concept of *Perfect Bayesian Equilibrium* (in short, *PBE*). We will focus on pure strategy equilibria.

3 Sincere Equilibrium in Period 2

In this section, we identify an equilibrium of the period 2 continuation game where the deputy reports truthfully (*sincere* equilibrium). Moreover, when the continuation equilibrium is sincere, we find that the principal will select whichever expert is perceived to be less biased as the period 2 leader. This result in turn will shape the outcome in period 1, including the selection of the period 1 leader.

Consider Stage T_{21} at the start of period 2. For any $k (= 1, 2,$ or, $\ell, d)$, let $\mu_k(P)$ denote the principal's posterior that expert k is unbiased at the interim stage T_{inter} conditional on the history till that stage. Sometimes P will be omitted from $\mu_k(P)$ as this posterior will be relevant for the principal's decisions only.

Next we define *efficient project implementation* in period t, $t = 1, 2$.

Definition 1 *A project implementation rule is said to be efficient if the rule* (a) *implements a project according to the more talented* λ*-type agent's informative signal (i.e.,* $m = a_t$ *when her signal is* α*, and* $m = b_t$ *when her signal is* β*),* (b) *implements a project according to the less talented* ξ*-type agent's informative signal (i.e.,* $m = a_t$ *when her signal is* α*, and* $m = b_t$ *when her signal is* β*) when the talented agent receives no signal, and* (c) *the rule randomly implements either of the projects in case both agents receive no signal.*

We call this rule efficient since it maximizes expected returns from the project given the talent levels and the signals of the two agents. The proof, which is omitted, is straightforward given that the two states a_2 and b_2 have an equal chance of occurrence as specified in (1).

Definition 2 *A perfect Bayesian equilibrium of the continuation game starting in Stage* T_{22} *is* <u>sincere</u>*, if the period 2 deputy reports both her talent and signal truthfully, and the leader, if unbiased, uses the efficient project implementation rule.*

Proposition 1 ***(*Sincere equilibrium*)*** *Let Assumption 2 hold. Consider the continuation game starting in stage* T_{22} *of period 2, and suppose expert* i *is the leader. Then there exists a sincere equilibrium:*

1. *The deputy, i.e. expert j, $j \neq i$, reports both her talent and signal truthfully;*

2. *The leader selects **(a)** the efficient project if she is unbiased, whereas **(b)** she selects the project she favors, in case she is biased.*

The key intuition is that the deputy never gains from her bias, so she reports truthfully. Consequently, an unbiased leader makes her decision based on the deputy's information (which is truthful). Given Assumption 2, a biased leader however selects her pet project.

Given the sincere equilibrium, with an honest leader in period 2 the probability that the implemented project will be successful is:

$$X \equiv \left[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2 \cdot \frac{1}{2}\right], \tag{3}$$

Throughout we will substitute the symbol, X, when deriving the expression of period 2 payoffs.

In case period 2 leader is biased, then under the sincere equilibrium the expected period 2 payoff of the leader is $\zeta + \frac{1}{2}(sV_2)$, and that of the deputy is $\frac{sV_2}{2}$. Whereas if period 2 leader is honest, then her expected payoff and that of the deputy are identical, equal to $X(sV_2)$.

Given Proposition 1, we show that the principal makes the leadership decision based on expected bias alone. Intuitively, if the experts play a sincere equilibrium following the choice of leaders, then project selection will be efficient whenever the experts are unbiased, which is aligned with the principal's objective. The only issue is that the leader might be biased, which is minimized by choosing the expert with the least expected bias as the leader.

Proposition 2 *(**Value of honesty**) Consider the continuation game starting in stage T_{21} of period 2.*

1. *The principal's optimal decision is to choose expert i as the leader if and only if the principal considers her to be less biased, i.e., $\mu_i(P) \geq \mu_j(P)$.*

2. *Fix any period 1 strategy for the leader and the deputy such that both projects have a positive probability of both success and failure under all eventualities. Moreover, suppose the continuation game starting in Stage T_{22} involves playing the sincere equilibrium. It is optimal that the principal's regime-switching decision in period 2 only depends on project choice, and not on project success, or failure.*

Part 1 suggests that in period 1 the agents will place a premium on signalling their honesty whenever the continuation equilibrium is expected to be sincere. Further, part 2 shows that project success or failure does not affect what inference the principal draws regarding the honesty of the agents. It depends only on project choice (for the leader) and the recommendation (of the deputy).

Remark 1 *The period 2 continuation game also has a babbling equilibrium where the deputy makes a pooling recommendation, an unbiased leader makes a decision based on her own signal, and a biased leader implements her favored project. Moreover, given this equilibrium of the continuation game the principal makes the period 2 leadership choice based on both talents as well as expected biases of the two agents. We however focus on the sincere equilibrium since our objective is to examine incentives in period 1, and to that end we abstract from period 2 inefficiencies to the extent possible.*

4 Yesman-I: Less Talented but Ex-Ante more Honest Expert as the Leader

We shall focus on perfect Bayesian equilibria where, following our discussion in Remark 1, we assume that in period 2 the agents play a sincere equilibrium. Further, for tractability we focus on pure strategy equilibria. In what follows we shall use the term equilibrium to refer to PBE that satisfy both these properties. Given our motivation, we shall focus on equilibria where the deputy always submits a null report in period 1 irrespective of her signal, or bias type, referring to them as Yesman equilibria.[21]

We start by examining a scenario where the principal appoints the ex-ante more honest expert, agent 1, with a talent level of ξ as the period 1 leader, and agent 2 (the less honest) with a talent level of λ as her deputy. We shall call this leader-deputy combination the Yesman-I setup, and the corresponding equilibrium the Yesman-I equilibrium, where I is a mnemonic for the fact that agent 1 is the leader.

We first introduce some definitions that are useful for characterizing the leader's actions (recall that by equilibrium we mean PBE that satisfies the two properties delineated earlier).

Definition 3 *In any equilibrium we call the leader* **signal respecting** *if she implements* a_1 *when her own signal is* α*, and implements* b_1 *if her own signal is* β*.*

Definition 4 *In any equilibrium we call an honest a leader* **virtue signalling** *if, irrespective of her own signal, she implements the project that is not favored by her biased self.*[22]

Definition 5 *In any equilibrium we call a biased leader* **cut-throat** *if, irrespective of her own signal, she implements her favored project.*

We start by analyzing Yesman-I equilibria when the bias benefits are not too large.

Proposition 3 ***(*Leader signal respecting*)*** *Suppose Assumptions 1 and 2(a) hold, and* $\zeta \leq (2\xi - 1)sV_1$*. Then there exists an Yesman-I equilibrium where, in period 1, both unbiased and biased leaders are signal respecting, and in period 2, the principal does not induce a regime switch irrespective of project choice.*

The proof appears in Appendix C of Bag and Roy Chowdhury (2022).

Thus whenever $\zeta \leq (2\xi - 1)sV_1$, there always exists an Yesman-I equilibrium where there is no further inefficiency in period 1 project choice by the leader. In order to make it interesting, in what follows we therefore assume that bias payoffs are significant, formally $\zeta > (2\xi - 1)sV_1$. Further, for ease of exposition we impose the tie-breaking rule that in case of indifference the unbiased leader always implements the project which is not favored by her biased self.

Given that the deputy plays an uninformative yesman strategy, from Proposition 2(ii) earlier we can restrict attention to equilibria where the principal's regime-switching decision is contingent on

[21]In Section 7 we shall briefly examine equilibria which are in some sense the polar opposite, one where period 1 deputy truthfully reports all signals. A second reason why Yesman strategy of the deputy could be meaningful is its behavioral connection in hierarchical organizations where the subordinate may fear contradicting the boss especially when the latter's job may go to the former and thus there is an in-built tension due to conflicts of interest.

[22]The idea is same as 'political correctness'.

project implementation decision alone. We first examine project implementation decision when the principal's strategy involves regime switch if and only if a_1 is implemented.

Lemma 1 *Let Assumptions 1 and 2(a) hold, and suppose $\zeta > (2\xi - 1)sV_1$. Consider any Yesman-I equilibrium where, in period 2, there is regime switch if and only if a_1 is implemented in period 1.*

(i) *On observing a signal of either β, or $\emptyset$, the unbiased leader always implements b_1.*

(ii) *On observing a signal of α, the unbiased leader implements a_1 if and only if*

$$(2\xi - 1)V_1 \geq (1 - \nu_2)[X - \frac{1}{2}]V_2. \tag{4}$$

(iii) *On observing a signal of either α, or $\emptyset$, a biased leader prefers to implement a_1. On observing a signal β, a biased leader prefers to implement b_1 if and only if*

$$(2\xi - 1)V_1 \geq \nu_2[X - \frac{1}{2}]V_2. \tag{5}$$

Lemma 1 is quite intuitive. An unbiased leader of course has no incentive to go against a signal of β, because not only does doing so reduce her period 1 payoff, it leads to regime switch with a possibly biased agent 2 gaining leadership position. With a signal of α however there is a tradeoff: while respecting the signal improves period 1 payoffs, it also leads to a regime switch with a possibly biased leader, so such a choice imposes restrictions that are formalized in (4). Next consider a biased leader with a signal of α. As far as her bias benefits are concerned, she is indifferent between obtaining ζ now (by implementing a_1), or obtaining it later (by implementing b_1 so that she retains leadership and implements a_1 in period 2). Thus her decision is solely driven by her expected non-bias payoff. By implementing a_1 now, she ensures that period 1 project choice is optimal, though, with regime switch there is some chance of her period 2 non-bias payoff being affected if the period 2 leader is biased. Whereas if she implements b_1, then her project choice in both periods is necessarily sub-optimal, so that she implements a_1. If the signal is β, again there are tradeoffs formalized by (5).

This lemma allows us to formalize the notion of agent 2's degree of honesty, i.e. ν_2 being large, small, or intermediate, something that will play a role in Section 6 later. Note that (4) is more likely to hold if, keeping other parameters constant, ν_2 is large. Whereas (5) is more likely to hold if ν_2 is small. Following from this observation, we have:

Definition 6 (Degree of honesty for agent 2) *1. We say that agent 2 is* ex-ante *honest/unbiased, i.e. ν_2 is* large *whenever (4) holds, whereas (5) is violated.*

2. We say that agent 2 is ex-ante *biased, i.e. ν_2 is* small *whenever (4) is violated, whereas (5) holds.*

3. We say that agent 2 is ex-ante *biased relative to agent 1, i.e. ν_2 is small relative to ν_1, whenever $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} \geq \nu_2$.*[23]

[23]Note that $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} \geq \nu_2$ does not hold for $\nu_1 = \nu_2$, and requires ν_2 to be small compared to ν_1.

4. *We say that* ex ante*, agent 2 has an intermediate level of honesty, whenever* (4) *and* (5) *are both violated.*[24]

Of course, when we say ν_2 is large, small, or intermediate, this is after fixing the other relevant parameters. Thus, for example, agent 2's honesty level being large, or small, can alternatively be thought of as V_1/V_2 takes an intermediate value, etc.

Lemma 2 below allows us to rule out several possible regime-switching strategies by the principal.

Lemma 2 *Let Assumptions 1 and 2(a) hold, and suppose $\zeta > (2\xi - 1)sV_1$.*

(i) *There does not exist any Yesman-*I *equilibrium where there is* **necessarily** *regime switch irrespective of project choice.*

(ii) *If $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} < \nu_2$, then there does not exist any Yesman-*I *equilibrium where there is* **no** *regime switch irrespective of project choice.*

(iii) *If* (4) *holds, then there does not exist any Yesman-*I *equilibrium where there is regime switch if and only if b_1 is implemented.*

This helps us to focus only on three types of regime switching. Of these, triggering a switch when the leader implements her corrupt self's favorite project allows for more varied responses by the leader to her signal, depending on her bias type. Equilibrium strategies not only cater to efficiency considerations or corrupt implementation gains but also are motivated by the incentives of not losing leadership. The following proposition is a complete characterization of Yesman-I equilibrium.

Proposition 4 ***(*****Yesman**-I*)* *Suppose expert 1 is made the period 1 leader, and Assumptions 1 and 2(a) hold. Also, suppose $\zeta > (2\xi - 1)sV_1$.*

1. *Consider Yesman-*I *equilibria where, in period 2, the principal induces a regime switch if and only if a_1 is implemented:*

(A) (Agent 2 ex-ante honest: Unbiased leader signal respecting, biased leader cut-throat)*: If $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} < \nu_2$,* (4) *holds, and* (5) *is violated, then there exists an Yesman-*I *equilibrium where, in period 1, the unbiased leader is signal respecting, and the biased leader is cut-throat (see Fig. 1):*

$$\mathcal{I}_{1,1}(\xi) = \begin{cases} (N, \{\alpha\}) \to a_1 \\ (N, \{\beta, \emptyset\}) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to a_1 \end{cases}$$

(B) (Agent 2 has an intermediate level of honesty: Unbiased leader virtue signalling, biased leader cut-throat)*: If* (4) *and* (5) *are both violated, then there exists an Yesman-*I *equilibrium where, in period 1, the unbiased leader is virtue signalling, and the biased leader is*

[24] It seems natural to say that ν_2 is intermediate for the case when (4) and (5) both hold also. As Proposition 4 later shows however, an equilibrium where (4) and (5) both hold can only exist if one also has that ν_2 is small relative to ν_1. Consequently we refrain from interpreting (4) and (5) both holding as being ν_2 intermediate.

cut-throat (see Fig. 1):

$$\mathcal{I}_{1,1}(\xi) = \begin{cases} (N, \{\alpha, \beta, \emptyset\}) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to a_1 \end{cases}$$

(C) (Agent 2 ex-ante biased: Unbiased leader virtue signalling, biased leader signal respecting)*: If* (4) *is violated and* (5) *holds, then there exists an Yesman-*I *equilibrium where, in period 1, the unbiased leader is virtue signalling, and the biased leader is signal respecting (see Fig. 1):*

$$\mathcal{I}_{1,1}(\xi) = \begin{cases} (N, \{\alpha\}) \to a_1 \\ (N, \{\beta, \emptyset\}) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to a_1 \end{cases}$$

(D) *If* (4) *and* (5) *both hold, then, within the class where the principal induces a regime switch if and only if* a_1 *is implemented, no equilibrium exists.*

2. (ν_2 is small relative to ν_1: Unbiased leader signal respecting, biased leader cut-throat)*: Consider a candidate Yesman-*I *equilibrium where, in period 1, the unbiased leader is signal respecting, the biased leader is cut-throat, and, in period 2, the principal never changes the period 1 leader. This equilibrium exists if and only if* $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} \geq \nu_2$ *(see Figs. 1 and 2):*

$$\mathcal{I}_{1,1}(\xi) = \begin{cases} (N, \{\alpha\}) \to a_1 \\ (N, \{\beta, \emptyset\}) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to a_1 \end{cases}$$

3. (Complete pooling)*: Consider a candidate Yesman-*I *equilibrium where, in period 1, the leader implements* a_1 *irrespective of her bias, or signal, and there is regime switch if and only if* b_1 *is implemented. This equilibrium exists if and only if* (4) *is violated (see Fig. 2):*

$$\mathcal{I}_{1,1}(\xi) = (\{N, \neg N\}, \{\alpha, \beta, \emptyset\}) \to a_1.$$

4. *Apart from the equilibria described in Proposition 4.*1(A)*, 4.*1(B)*, 4.*1(C)*, 4.*2 *and 4.*3*, the only other equilibria that exist differ from one of these only in the principal's off-equilibrium beliefs.*

Among the different implementation strategies, Case 1.(B) is important for its potential for generating maximal period 1 inefficiency. Virtue signalling by an unbiased leader comes at a cost as she abstains from truthfully implementing project a_1 when her signal is α. This can happen when the probability of the deputy being biased, $1-\nu_2$, is relatively large, so worries of period 2 inefficiency wins over concerns of period 1 inefficiency, and especially so because V_2 exceeds V_1. At the same time, the biased leader will compromise on period 1 inefficiency in order to reap the gains of her corrupt spoils, ζ, when the deputy's chance of being honest, ν_2, is sufficiently large; expert 2, who will be in charge of project 2, is then quite likely to choose informationally efficient project. These two conditions, together, imply that ν_2 is neither too small nor too large, as illustrated in Fig. 1 with

$0.15 < \nu_2 < 0.85$ when $V_2 = 4750$ and $V_1 = 1000$, along with other parameter specifications. For this specific illustration, the set of (ν_1, ν_2) supporting the equilibrium 1.(B) is quite large.

Case 3 (complete pooling) is perverse in that the leader whose biased self is known to favor project a_1 will be replaced provided she implements project b_1 instead. Not only the switching rule is unappealing, it also yields the worst payoff for the principal as we show in Lemma 3. Even with little commitment power organizations should be able to rule out this replacement rule. We ignore this equilibrium from further analysis, relegating it to Appendix C, Bag and Roy Chowdhury (2022).

Virtue signalling and signal respecting (Case 1.(C)) are likely to arise when ν_2 is small with no restriction on ν_1. The reason is, the unbiased leader is willing to compromise on period 1 inefficiency (loss of sV_1) to guarantee that relatively high period 2 payoff (sV_2) is not lost due to expert 2 being highly likely to be biased who would then implement informationally wrong project b_1 under α signal. For the biased leader, the intuition for signal respecting behavior appears following Lemma 1.

The intuitions for the remaining cases can be similarly understood and easily mapped to Figs. 1 and 2.

To recap, in Figs. 1 and 2 we present a partition of the space $\{(\nu_1, \nu_2) \mid \text{Yesman-I}, \nu_1 \geq \nu_2\}$ into different types of equilibria. Note here that some regions will have multiple equilibria, so the principal needs to compare the payoffs across these equilibria for his leadership delegation decision in Section 6.

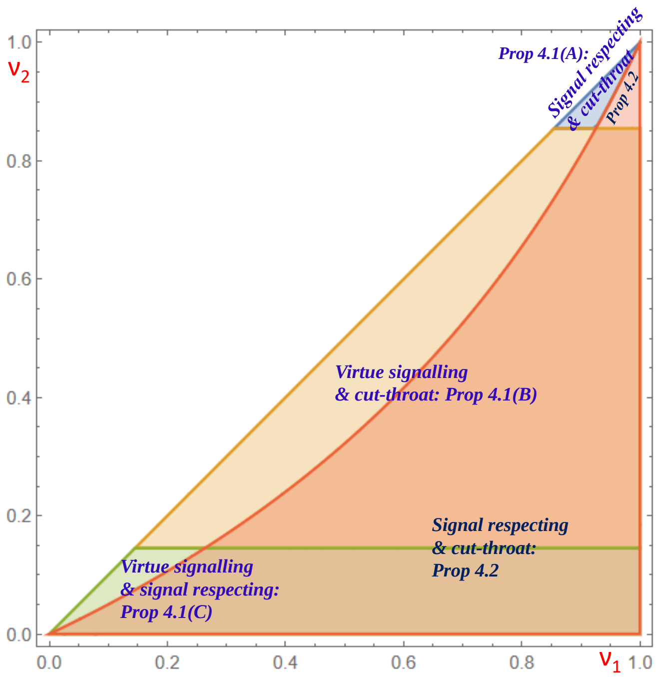


Figure 1: Yesman-I equilibria in Proposition 4.1(A) − (C) and Proposition 4.2 plotted against (ν_1, ν_2), fixing $\xi = 0.6, \lambda = 0.8, \epsilon = 0.05, s = 0.1, V_1 = 1000, V_2 = 4750, \zeta \geq 340$. Summary: Multiple equilibria

■ **Principal's payoffs.** We now turn to the determination of the present value of the principal's expected payoffs over the two periods (henceforth principal's surplus) for the equilibria in Proposition 4. Here we report only the algebraic expressions.

▶ The principal's surplus in the equilibrium described in Proposition 4.1(A) is:

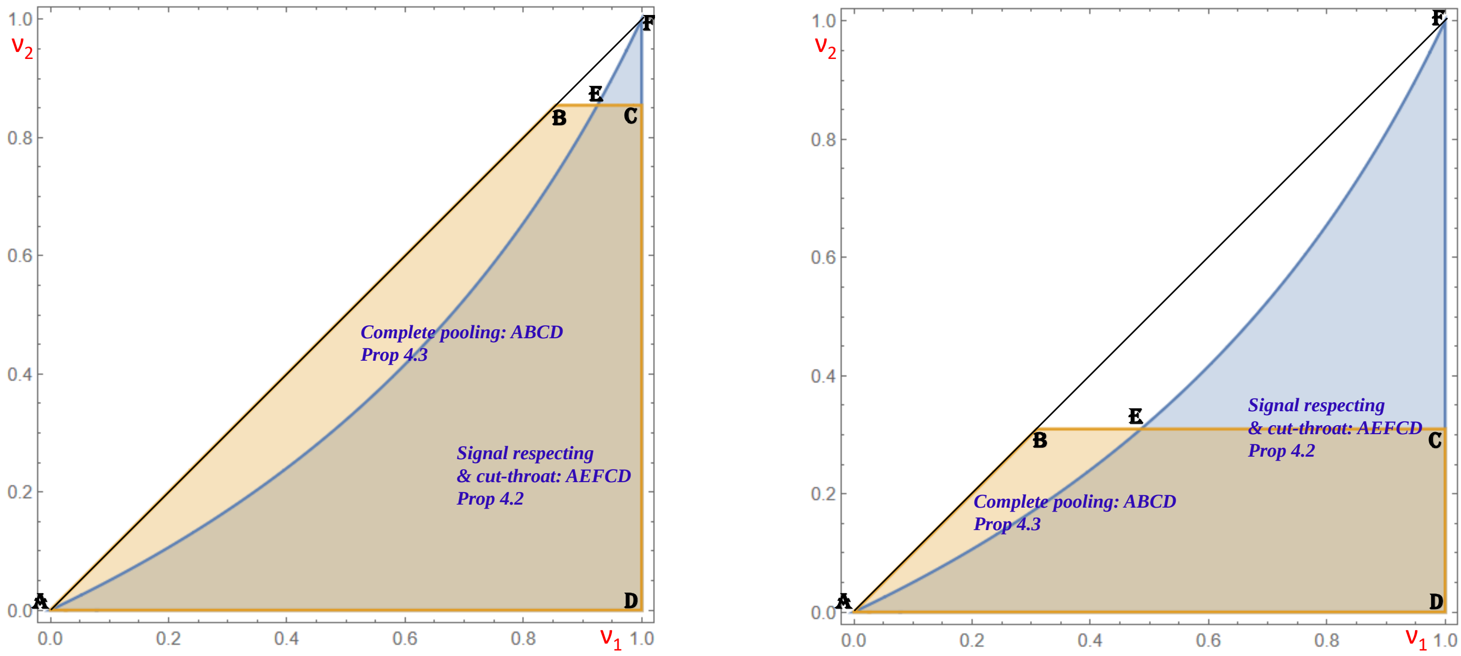


Figure 2: Yesman-I equilibria (Proposition 4.2/3): First panel, $\xi = 0.6, \lambda = 0.8, \epsilon = 0.05, s = 0.1, V_1 = 1000, V_2 = 4750, \zeta \geq 340$. Second panel, $\xi = 0.6, \lambda = 0.8, \epsilon = 0.05, s = 0.1, V_1 = 1000, V_2 = 1000, \zeta \geq 100$.

$$
\begin{aligned}
&\nu_1 \Bigg[\frac{1}{2}\{(1-\epsilon)\xi\}\Big\{ s_P V_1 + \Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big) s_P V_2 \Big\} + \frac{1}{2}\{\epsilon + (1-\epsilon)(1-\xi)\}\Big\{ X(s_P V_2) \Big\} \\
&\quad + \frac{1}{2}\{\epsilon + (1-\epsilon)\xi\}\Big\{ s_p V_1 + X(s_P V_2) \Big\} + \frac{1}{2}\{(1-\epsilon)(1-\xi)\}\Big\{ \Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big) s_P V_2 \Big\} \Bigg] \\
&+ (1-\nu_1)\Big[\frac{1}{2}(s_P V_1) + \Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big) s_P V_2 \Big] \\
= &\ \frac{1}{2}\big[\nu_1\{\epsilon + 2(1-\epsilon)\xi\} + (1-\nu_1)\big](s_P V_1) \\
&\quad + \big[\frac{1}{2}\nu_1(1-\epsilon) + (1-\nu_1)\big]\Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)(s_P V_2) + \frac{1}{2}\nu_1(1+\epsilon)X(s_P V_2).
\end{aligned} \tag{6}
$$

▶ The principal's surplus in the equilibrium described in Proposition 4.1(B) is:

$$
\frac{1}{2}(s_P V_1) + \nu_1 X(s_P V_2) + (1-\nu_1)\big[\nu_2 X + (1-\nu_2)\frac{1}{2}\big](s_P V_2). \tag{7}
$$

▶ The principal's surplus in the equilibrium described in Proposition 4.1(C) is:

$$
\begin{aligned}
&\nu_1 \Big[\frac{1}{2}(s_P V_1) + X(s_P V_2) \Big] \\
&+ (1-\nu_1)\Bigg[\frac{1}{2}\{(1-\epsilon)\xi\}\Big\{ s_P V_1 + \frac{1}{2} s_P V_2 \Big\} \\
&\qquad + \frac{1}{2}\{\epsilon + (1-\epsilon)(1-\xi)\}\Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big) s_P V_2 + \frac{1}{2}\{(1-\epsilon)(1-\xi)\}\Big\{ \frac{1}{2} s_P V_2 \Big\} \\
&\qquad + \frac{1}{2}\{\epsilon + (1-\epsilon)\xi\}\Big\{ s_P V_1 + \Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big) s_P V_2 \Big\} \Bigg]
\end{aligned}
$$

$$= \frac{1}{2}\Big[\nu_1 + (1-\nu_1)\{\epsilon + 2(1-\epsilon)\xi\}\Big](s_P V_1)$$
$$+ \Big[\nu_1 X + (1-\nu_1)\frac{1}{2}\{\frac{1}{2}(1-\epsilon)\}\Big](s_P V_2) + (1-\nu_1)\Big[\frac{1}{2}(1+\epsilon)\Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)\Big](s_P V_2). \quad (8)$$

▶ The principal's surplus in the equilibrium described in Proposition 4.2 is:

$$\Big[\nu_1 \frac{1}{2}\{(1-\epsilon)\xi + \epsilon + (1-\epsilon)\xi\} + (1-\nu_1)\frac{1}{2}\Big](s_P V_1) + \Big[\nu_1 X + (1-\nu_1)\frac{1}{2}\Big](s_P V_2). \quad (9)$$

The following surplus comparisons, proved in the working version of the paper, will be useful in our subsequent analysis.[25] The proof is omitted but appears in Bag and Roy Chowdhury (2022).

***Lemma 3* (Complete pooling: the worst)** *In the case of another equilibrium co-existing with the complete pooling equilibrium in Proposition 4.3, each of the principal's surplus expressions* (7), (8), *and* (9) *is strictly greater than the surplus in the complete pooling equilibrium.*

5 Yesman-II: More Talented but more Biased Expert as the Leader

Consider the continuation game where the principal appoints agent 2 as the period 1 leader, and agent 1 as her deputy. We call this the Yesman-II scenario, where II stands for agent 2 as the leader. We shall examine Yesman equilibria of this game, calling such equilibria Yesman-II.

As in the preceding section, one can demonstrate that when bias benefits are small, formally $\zeta \leq (2\lambda - 1)sV_1$, there exists an Yesman-II equilibrium where the leader is signal respecting, so that there is no inefficiency in period 1 project choice by the leader.[26] In what follows we therefore assume that bias payoffs are significant, formally $\zeta > (2\lambda - 1)sV_1$ which, recall, is our Assumption 2(b).

Recall that the pet project of the biased agent 2 (the period 1 leader) is b_1. We start by showing that in any Yesman-II equilibrium the principal's strategy must involve a regime switch if and only if period 1 leader implements b_1.

***Lemma 4* (Ruling out specific regime-switching)** *Let Assumptions 1 and 2 both hold.*

(i) *There does not exist any Yesman-*II *equilibrium where either (a) there is no regime switch irrespective of project outcome, or (b) there is necessarily regime switch irrespective of project outcome.*

(ii) *There does not exist any Yesman-*II *equilibrium where there is regime switch if and only if a_1 is implemented.*

[25] It may be noted that leaving aside the complete pooling equilibrium, there can be multiple equilibria reported in Proposition 4.1(B), (C) and 4.2, as illustrated in Figs. 1 and 2. It can be shown that the surplus differences (7)−(9) and (8)−(9) can be either positive or negative.

[26] By signal respecting, a talented corrupt leader will receive at most $\lambda(sV_1)$ + continuation payoff (call it Z_2) (this payoff is likely to be depressed due to the loss of leadership). By not respecting the signal she receives $\zeta + (1-\lambda)(sV_1) +$ a different continuation payoff (call it Z_2'). So if the first payoff (weakly) exceeds the second payoff, then there will never be any period 1 inefficiency. For this not to happen, it must necessarily be that $\zeta + (1-\lambda)(sV_1) + Z_2' > \lambda(sV_1) + Z_2$, which can be rewritten as $\zeta > (2\lambda - 1)(sV_1) + (Z_2 - Z_2')$. But given that $Z_2' - Z_2 \geq 0$, the necessary condition to rule out 'no period 1 inefficiency' is $\zeta > (2\lambda - 1)sV_1$.

Given this lemma, we next characterize the set of Yesman-II equilibrium where there is regime switch if and only if b_1 is implemented. We start by examining project implementation decision under such an equilibrium.

Lemma 5 *(* **Project choices under suspected bias** *) Let Assumptions 1 and 2 both hold. Consider Yesman-*II *equilibrium where, in period 2, there is regime switch if and only if b_1 is implemented in period 1.*

(i) *On observing a signal of either α, or $\emptyset$, the unbiased leader always implements a_1.*

(ii) *On observing a signal of β, the unbiased leader implements b_1 if and only if*

$$(2\lambda - 1)V_1 \geq (1 - \nu_1)\big[X - \frac{1}{2}\big]V_2. \tag{10}$$

(iii) *On observing a signal of either β, or $\emptyset$, a biased leader prefers to implement b_1.*

On observing a signal α, a biased leader prefers to implement a_1 if and only if

$$(2\lambda - 1)V_1 \geq \nu_1\big[X - \frac{1}{2}\big]V_2. \tag{11}$$

The intuition is very similar to that of Lemma 1 earlier, and hence omitted. Moreover, we can use this lemma to formalize the notion of agent 1's degree of honesty.

Definition 7 (Degree of honesty for agent 1) *1. We say that agent 1 is* ex-ante *honest, i.e. ν_1 is* large *whenever (10) holds, whereas (11) is violated.*

2. We say that agent 1 is ex-ante *biased, i.e. ν_1 is* small *whenever (4) is violated, whereas (5) holds.*

3. We say that ex ante, *agent 1 has an intermediate level of honesty, whenever either both (10) and (11) hold, or both are violated.*

Proposition 5 below is the central result of this section.

Proposition 5 *(* **Equilibrium under talent as priority** *) Suppose expert 2 is made the period 1 leader, and Assumptions 1 and 2 both hold.*

1. (Agent 1 ex-ante honest: Unbiased leader signal respecting, biased leader cut-throat): *If (10) holds, and (11) is violated, then there exists an Yesman-*II *equilibrium where, in period 1, the unbiased leader is signal respecting, the biased leader is cut-throat, i.e.,*

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \emptyset\}) \to a_1 \\ (N, \beta) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to b_1 \end{cases}$$

and there is regime switch if and only if b_1 is implemented.

2. (Agent 1 has an intermediate level of honesty: Unbiased leader virtue signalling, biased leader cut-throat): *If* (10) *and* (11) *are both violated, then there exists an Yesman*-II *equilibrium where, in period 1, the unbiased leader is virtue signalling, the biased leader is cut-throat, i.e.,*

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \beta, \emptyset\}) \to a_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to b_1 \end{cases}$$

and there is regime switch if and only if b_1 *is implemented.*

3. (Agent 1 ex-ante biased: Unbiased leader virtue signalling, biased leader signal respecting): *If* $\frac{2\nu_2}{2\nu_2+(1-\nu_2)(1-\epsilon)} < \nu_1$, (10) *is violated, and* (11) *holds, then there exists an Yesman*-II *equilibrium where, in period 1, the unbiased leader is virtue signalling, the biased leader is signal respecting, i.e.,*

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \beta, \emptyset\}) \to a_1 \\ (\neg N, \{\beta, \emptyset\}) \to b_1 \\ (\neg N, \{\alpha\}) \to a_1 \end{cases}$$

and there is regime switch if and only if b_1 *is implemented.*

4. (Agent 1 has an intermediate level of honesty: Leader signal respecting): *There exists a unique* $\hat{\epsilon} < 1$ *such that an Yesman*-II *equilibrium satisfying conditions* (10) *and* (11) *exists if and only if* $\epsilon \geq \hat{\epsilon}$. *Further, this equilibrium exists for* __*all*__ $\epsilon \geq 0$ *if* $V_1 \geq \max\{\nu_1, 1-\nu_1\} \cdot \frac{V_2}{2}$. *In this equilibrium, in period 1, both unbiased and biased leaders are signal respecting, i.e.,*

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (\{N, \neg N\}, \{\alpha\}) \to a_1 \\ (\{N, \neg N\}, \{\beta\}) \to b_1 \\ (\neg N, \{\emptyset\}) \to b_1 \\ (N, \{\emptyset\}) \to a_1 \end{cases}$$

and there is regime switch if and only if b_1 *is implemented.*

5. *Apart from the equilibria described in Proposition 5*.1*, 5*.2*, 5*.3*, and 5*.4*, the only other equilibria that exist differ from one of these only in the principal's off-the-equilibrium beliefs.*

The difference here from Proposition 4 is that we have a <u>unique</u> equilibrium, for any given set of parameters, although the nature of equilibrium would still vary across different ranges of parameters with combinations of period 1 inefficiencies and screening (or non-screening) of biased types. There are two types inefficiencies in project choice: (i) unbiased leader engages in virtue signalling, (ii) biased leader implements corrupt project despite contrary signal, with the latter encouraging the former because otherwise the principal would change leadership when questionable project (b_1) is implemented. Three distinct regions emerge as possible equilibrium that are displayed in Fig. 3. Given the assumption that $\nu_1 > \nu_2$, our main interest is in the area below the 45-degree line. Note that the middle region (part (2)) is where the leader makes both types of inefficient implementation

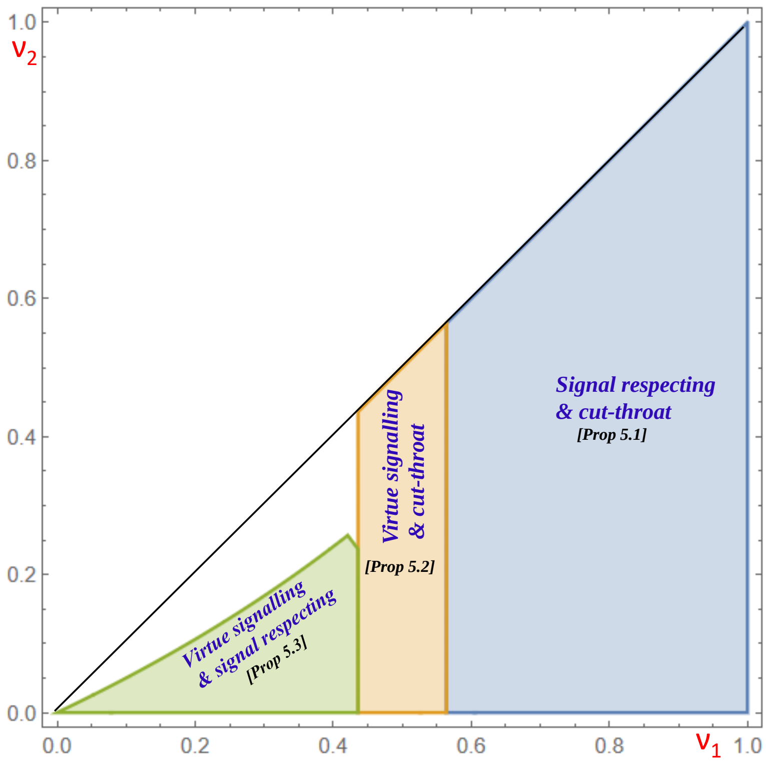


Figure 3: Yesman-II equilibrium region plotted against (ν_1, ν_2), fixing $\xi = 0.6, \lambda = 0.8, \epsilon = 0.05, s = 0.1, V_1 = 1000, V_2 = 4750, \zeta \geq 350$. Virtue signalling prompts the unbiased self of the less perceived-to-be honest but skilled expert (agent 2) to always implement a_1 irrespective of signal, thus improving her perception for honesty (parts (1) & (2) of Proposition 5). Although not displayed in this Figure, for these same parameter values, the leader is always signal respecting (Proposition 5.4) for all (ν_1, ν_2) in the entire lower triangle if $\epsilon \geq 0.7$. An alternative plot (not reported here) shows that if $V_1 = 1000$ were replaced by $V_1 = 1300$, the middle space occupied by Proposition 5.2 will be exactly replaced by Proposition 5.4.

– unbiased leader engages in virtue signaling (or what Morris (2006) called political correctness) and biased leader chooses her corrupt project. This region is concentrated around ν_1 in the middle: if ν_1 is high then an unbiased leader would rather implement informatively efficient project in period 1 because in the event of leadership change period 2 decisions are more likely to be efficient; if ν_1 is low, the unbiased leader would engage in virtue signalling (so long as V_1 is relatively much less than V_2) because she worries that change of leadership will likely put a corrupt expert (agent 1) in charge, thus inflicting a substantial loss of project efficiency. For the biased leader (agent 2), if ν_1 is high (indicated in the blue and brown regions of Fig. 3) then she always implements her favorite project (b_1) in period 1 irrespective of signal realization and reap the corrupt gains because with the resulting change of leadership to agent 1 there is a high chance of ex-ante efficient implementation in period 2. If however ν_1 is small (green region), inefficient implementation (b_1 when signal is α) is costly both in period 1 as well as period 2, with period 2 cost coming from a high chance of corrupt implementation by the new leader agent 1. Finally, when ϵ is sufficiently high (see remark in the descriptive caption of Fig. 3, $\epsilon \geq 0.7$), both types of leader are signal respecting for most (or all) of (ν_1, ν_2) (part 4 of the proposition), because the leader benefit from efficient implementation and the biased type, by implementing b_1 when signal is β, enjoys additionally the corrupt gains; the concern for period 2 inefficiency, should leadership be transferred to agent 1, does not bother agent 2 because the chances of informative implementation in period 2 is not that high (high ϵ).

■ **Principal's payoffs.** We now turn to the determination of principal's surplus for the different equilibria in Proposition 5.

▶ The principal's surplus in the equilibrium described in Proposition 5.1 equals:

$$\begin{aligned}
&\nu_2\Bigg[\frac{1}{2}\{\epsilon+(1-\epsilon)\lambda\}\Big\{s_P V_1 + X(s_P V_2)\Big\} + \frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\Big\{\Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\} \\
&\quad + \frac{1}{2}\{\epsilon+(1-\epsilon)(1-\lambda)\}\Big\{X(s_P V_2)\Big\} + \frac{1}{2}\{(1-\epsilon)\lambda\}\Big\{s_p V_1 + \Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\}\Bigg] \\
&+ (1-\nu_2)\Bigg[\frac{1}{2}(s_P V_1) + \Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Bigg].
\end{aligned} \tag{12}$$

▶ The principal's surplus in the equilibrium described in Proposition 5.2 equals:

$$\frac{1}{2}(s_P V_1) + \nu_2\big[X(s_P V_2)\big] + (1-\nu_2)\Bigg[\Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)(s_P V_2)\Bigg]. \tag{13}$$

▶ The principal's surplus in the equilibrium described in Proposition 5.3 equals:

$$\begin{aligned}
&\nu_2\Bigg[\frac{1}{2}(s_P V_1) + X(s_P V_2)\Bigg] \\
&+ (1-\nu_2)\Bigg[\frac{1}{2}\{\epsilon+(1-\epsilon)\lambda\}\Big\{s_P V_1 + \Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\} + \frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\Big\{\frac{1}{2}s_P V_2\Big\} \\
&\qquad + \frac{1}{2}\{\epsilon+(1-\epsilon)(1-\lambda)\}\Big\{\Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\} + \frac{1}{2}\{(1-\epsilon)\lambda\}\Big\{s_P V_1 + \frac{1}{2}s_P V_2\Big\}\Bigg].
\end{aligned} \tag{14}$$

▶ The principal's surplus in the equilibrium described in Proposition 5.4 equals:

$$\begin{aligned}
&\Bigg[\frac{1}{2}\{(1-\epsilon)\lambda\}\Big\{s_P V_1 + \Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)s_P V_2\Big\} \\
&\quad + \frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\Big\{\Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\} + \frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\Big\{\Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)s_P V_2\Big\} \\
&\quad + \frac{1}{2}\{(1-\epsilon)\lambda\}\Big\{s_p V_1 + \Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\}\Bigg] \\
&+ (1-\nu_2)\Bigg[\epsilon\Big\{\frac{1}{2}s_P V_1 + \Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\}\Bigg] + \nu_2\Bigg[\epsilon\Big\{\frac{1}{2}s_P V_1 + \Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)s_P V_2\Big\}\Bigg].
\end{aligned} \tag{15}$$

6 Who to Delegate Period 1 Authority to: Talented or More Honest Expert?

In this section, we finally turn to one of the key questions of this paper: the principal's leadership choice in period 1.

Given that we are interested in scenarios where agent 1 is *ex ante* more honest but less talented relative to agent 2, we shall focus on parameter values such that the *ex-ante* honesty differential $\nu_1 - \nu_2$ is *not too small*,[27] drawing on definitions 6 and 7 to formalize this idea. To be precise we focus on

[27] Though for completeness in Section 6.1 we shall also report on the case where the honesty differential is small.

four parameter configurations, (i) ν_1 is large and ν_2 is small, (ii) ν_1 is large, and ν_2 is small relative to ν_1, (iii) ν_1 is large, and ν_2 is intermediate, and (iv) ν_1 is intermediate, and ν_2 is small.[28]

For ease of exposition, going forward we shall write talent wins (resp. honesty wins) to imply that the principal selects agent 2 (resp. agent 1) as the period 1 leader. Our key result is that talent wins whenever ν_1 is large, and ν_2 is not too large. However, honesty may win whenever ν_1 is intermediate, and ν_2 is small. As we argue in greater details later, these results follow from the interaction of two factors, the possibility of virtue signalling, and the negative correlation between period 1 and period 2 surplus.

Recall from Lemma 3 that the complete pooling equilibrium, in which a period 1 leader is *retained* if and only if she implements the **favorite** project of her biased self, leads to a lower surplus compared to any other equilibrium that may exist. Further, changing leadership for choosing a project that won't privately benefit even a corrupt leader, which is critical for the complete pooling equilibrium of Proposition 4, is a perverse rule. For both these reasons it requires little persuasion by the governing board in any organization to prohibit such regime-switching rules. Further, given that project implementation is verifiable, such a commitment will be credible.[29]

Assumption 3 *The principal commits to not implementing a regime-switching rule such that period 1 leader is retained if and only if she implements the project favored by her biased self.*

The principal's choice of leadership will often depend on the *worth* of the two agents. The worth is a composite indicator determined by the level of *net talent*, $\tau_i - \frac{1}{2}(1-\epsilon)$,[30] and the equilibrium implementation strategy if appointed as the leader in period 1.

Definition 8 *The* **worth** *of an expert* i*, call it* $W(\nu_i, \tau_i)$*, is the ex-ante expected accuracy of expert* i*'s truthful implementation of her signal over and above the prior in period 1, recognizing that an expert observes an informative signal with probability* $1-\epsilon$*:*

$$W(\nu_i, \tau_i) \equiv \begin{cases} \nu_i(\tau_i - \frac{1}{2}(1-\epsilon)) & \text{when the unbiased type is signal respecting,} \\ (1-\nu_i)(\tau_i - \frac{1}{2}(1-\epsilon)) & \text{when the biased type is signal respecting.} \end{cases} \tag{16}$$

Proposition 6 below shows that talent wins whenever ν_1 is large, and ν_2 is either intermediate, or small (either absolutely, or relative to ν_1).

Proposition 6 (*Talent dominates*) *Let Assumptions 1, 2 and 3 all hold. Suppose ν_1 is large so that under the Yesman-II scenario the equilibrium in Proposition 5.1 obtains with the surplus given by (12).*

[28]From definitions 6 and 7 recall that ν_1 is large means that under the Yesman-II scenario the equilibrium in Proposition 5.1 obtains, that ν_2 is small means that under the Yesman-I scenario the equilibrium in Proposition 4.1(C) obtains, ν_2 is intermediate means that under the Yesman-I scenario the equilibrium in Proposition 4.1(B) arises, ν_2 is small relative to ν_1 means that under the Yesman-I scenario the equilibrium in Proposition 4.2 obtains, and that ν_1 is intermediate means that under the Yesman-II scenario either the equilibrium in Proposition 5.2, or the one in Proposition 5.4 obtains.

[29]Tirole (2006) argues that it is the role of the board of a firm to provide this kind of broad policy guidance.

[30]Recall that $\tau_1 = \xi(1-\epsilon)$, and $\tau_2 = \lambda(1-\epsilon)$.

1. *Suppose* ν_2 *is small, i.e. under the Yesman-*I *scenario, the equilibrium in Proposition 4.*1(C) *obtains when the surplus is* (8)*. The principal then appoints the talented but less honest expert 2 as the leader in period 1 if the worth of the talented agent exceeds the worth of the honest agent, i.e.*

$$\nu_2(\lambda - \frac{1}{2}) - (1 - \nu_1)(\xi - \frac{1}{2}) \geq 0,$$

where the honest agent's worth reflects that only her biased type will be signal respecting. Otherwise, the principal appoints the more honest but less talented agent, expert 1, as the period 1 leader. See the left panel Fig. 4.

2. *Suppose* ν_2 *is small relative to* ν_1*, i.e. under the Yesman-*I *scenario the equilibrium in Proposition 4.2 obtains with the surplus given by* (9)*. The principal then appoints the talented but less honest expert 2 as the leader in period 1:*

$$\begin{aligned}(12) - (9) = (1-\epsilon)\big[\nu_2(2\lambda - 1) - \nu_1(2\xi - 1)\big]\frac{1}{2}(s_P V_1) \\ + \frac{1}{2}\nu_2 X (s_P V_2) + \frac{1}{2}\nu_2\epsilon(1-\nu_1)\big[X - \frac{1}{2}\big](s_P V_2) + \frac{1}{2}\big(\nu_1 X + (1-\nu_1)\frac{1}{2}\big)\frac{1}{2}(s_P V_2) > 0.\end{aligned}$$

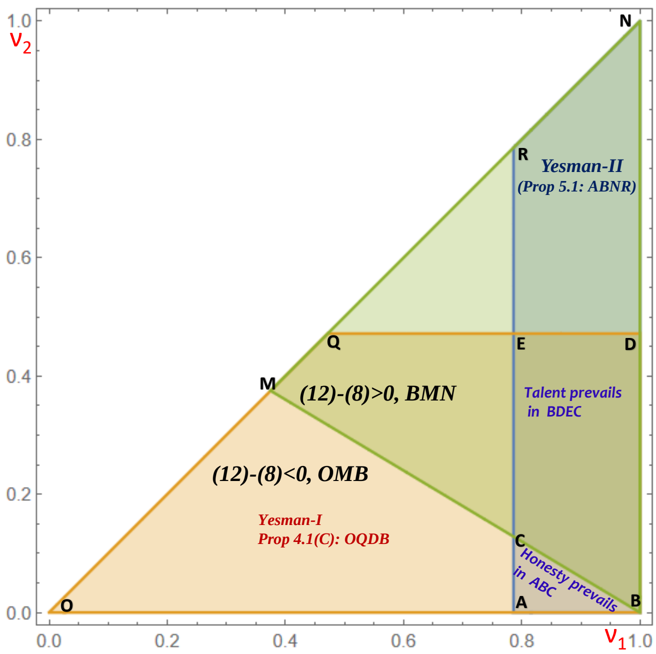
See the right-hand panel Fig. 4.

3. *Suppose* ν_2 *is intermediate, i.e. the Yesman-*I *equilibrium in Proposition 4.*1(B) *arises, with the principal's payoff given by* (7)*. Then the principal selects the talented expert as the period 1 leader (see Fig. 5):*

$$(12) - (7) = \frac{1}{2}\big\{\nu_2(1-\epsilon)(2\lambda - 1)\big\}(s_P V_1) + (1-\nu_1)\nu_2\frac{1}{2}(1-\epsilon)\big[1 - X\big](s_P V_2) > 0.$$

Turning to the intuition behind Proposition 6, for ease of exposition we again assume $\epsilon = 0$. Consider Proposition 6.1, when ν_2 is small. Under the Yesman-I equilibrium in Proposition 4.1(C), the unbiased leader is virtue signalling and the biased leader is signal respecting, so that her worth is $(1-\nu_1)(\xi - \frac{1}{2})(1-\epsilon)$, whereas under the Yesman-II equilibrium in Proposition 5.1, the unbiased leader is signal respecting and the biased leader is cut-throat, so that her worth is $\nu_2(\lambda - \frac{1}{2})(1-\epsilon)$. Hence period 1 project choice is more efficient under agent 2 leadership relative to that under agent 1 leadership iff the worth of agent 2 exceeds that of agent 1, i.e. $\nu_2(\lambda - \frac{1}{2}) - (1-\nu_1)(\xi - \frac{1}{2}) > 0$. What about period 2? Here the screening efficiency is the same irrespective of who is the period 1 leader: the probability that the second period leader is unbiased is $\frac{1}{2}\nu_2 + \frac{1}{2}\nu_2\nu_1 + (1-\nu_2)\nu_1$ under Yesman-II, and it is $\nu_1 + \frac{1}{2}(1-\nu_1)\nu_2$ under Yesman-I, the two probabilities being identical (a similar argument applies for $\epsilon > 0$). Thus overall efficiency is driven by period 1 considerations alone, hence the result.

Next, suppose that the honesty level of agent 2 is low relative to ν_1. Then under the Yesman-I scenario we have the equilibrium in Proposition 4.2 where the unbiased leader is signal respecting, whereas the biased leader is cut-throat. In this case, it is not clear whether the period 1 surplus is higher under Yesman-I, or Yesman-II. This is because with the honesty level of agent 2 being low, under the Yesman-II equilibrium the leader is unlikely to be honest, which is the only case when she is signal respecting. However, screening is more efficient under Yesman-II: the second period leader is unbiased with probability $\nu_1 + \frac{1}{2}(1-\nu_1)\nu_2$ under Yesman-II but it is ν_1 under Yesman-I (set $\epsilon = 0$ in



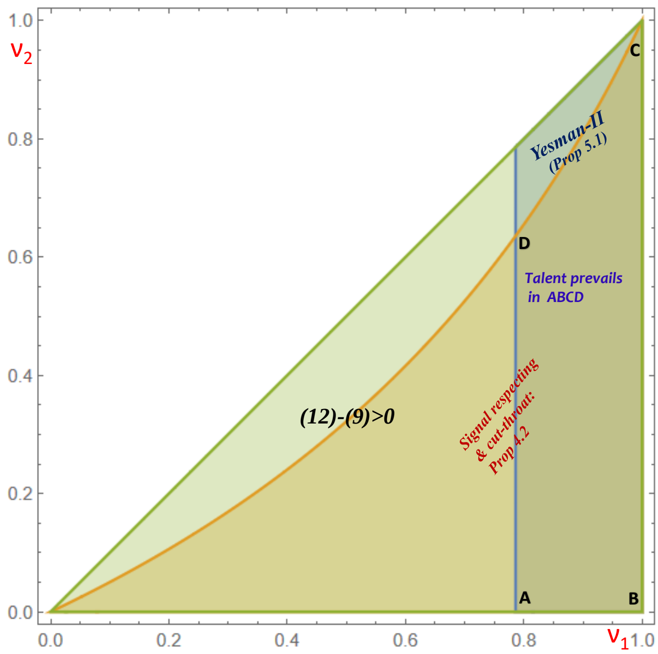


Figure 4: **[Left panel]** Region BDEC is where talent dominates honesty, and ABC is where honesty dominates [Proposition 6.1], fixing $\xi = 0.65, \lambda = 0.75, \epsilon = 0.05, s = 0.1, V_1 = 1000, V_2 = 2600, \zeta \geq 200$. **[Right panel]** Region ABCD is where talent dominates honesty [Proposition 6.2], fixing $\xi = 0.65, \lambda = 0.75, \epsilon = 0.05, s = 0.1, V_1 = 1000, V_2 = 2600, \zeta \geq 200$.

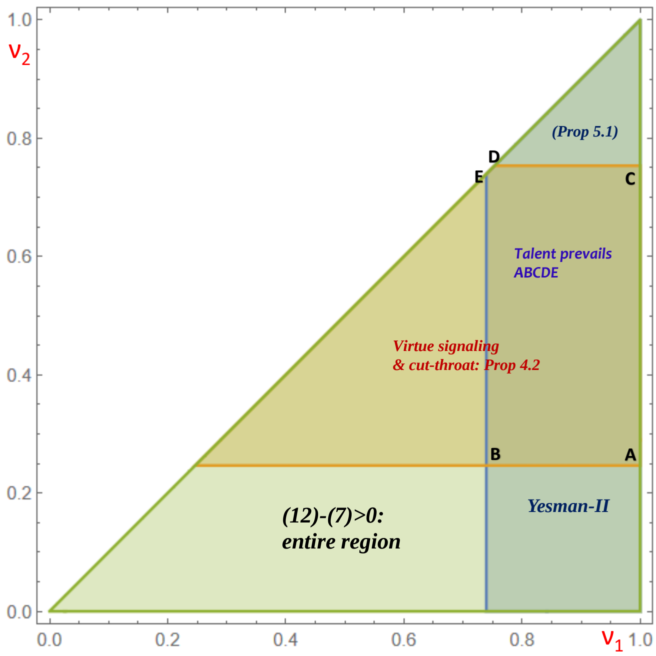


Figure 5: Region ABCDE is where talent dominates honesty [Proposition 6.3], fixing $\xi = 0.6, \lambda = 0.8, \epsilon = 0.05, s = 0.1, V_1 = 1000, V_2 = 2800, \zeta \geq 200$.

the first component of the second term in (9) and ignore λ). Thus there is an apparent tradeoff here. However, given that the biased leader is cut-throat, (11) is violated so that period 2 project return is relatively large compared to that in period 1. Thus the screening effect, which comes to fruition in period 2, dominates, hence the result.

Next consider the intuition for Proposition 6.3 when ν_2 is at an intermediate level. Under the Yesman-I equilibrium in Proposition 4.1(B), the unbiased leader is virtue signalling, whereas the biased leader is cut-throat, so that both kinds of leaders ignore their signal. In contrast, under Yesman-II in Proposition 5.1, the unbiased leader is signal respecting. Thus period 1 surplus is higher under the Yesman-II equilibrium. Period 2 surplus is, however, lower under the Yesman-II equilibrium since period 2 leader is going to be biased with probability $\frac{(1-\nu_1)(2-\nu_2)}{2}$, whereas under the Yesman-I equilibrium the period 2 leader is necessarily honest. The period 2 surplus effect is however small because of two reasons. First, period 1 payoffs are not too small (given that under the Yesman-II equilibrium the honest leader is signal-respecting), and second, in case talent wins the signal received by an honest agent 2 is not lost, which is an important driver of period 1 surplus given that ν_2 is intermediate.

Next suppose ν_1 is intermediate, and ν_2 is small. Note that ν_1 being intermediate is consistent with the Yesman-II equilibrium in Proposition 5.2, as well as that in Proposition 5.4. We next demonstrate that honesty wins whenever the Yesman-II equilibrium in Proposition 5.2 obtains. Otherwise, either talent or honesty may win.

Proposition 7 ***(*ν_1 intermediate, ν_2 small*)*** *Let Assumptions 1, 2 and 3 all hold. Suppose ν_2 is small, i.e. Yesman-*I *equilibrium in Proposition 4.*1(C) *exists, with the principal's payoff given by* (8).

1. *Honesty dominates: ν_1 is intermediate in the sense that the Yesman-*II *equilibrium in Proposition 5.*2 *obtains with the corresponding principal payoff given by* (13). *Then the principal selects expert 1, who is perceived to be more honest, as the period 1 leader:*

$$(8) - (13) = (1-\nu_1)(1-\epsilon)[\xi - \frac{1}{2}](s_P V_1) - \frac{1}{2}(1-\epsilon)(1-\nu_1)\nu_2[X - \frac{1}{2}](s_P V_2) > 0.$$

See the first panel of Fig. 6.

2. *Either honesty, or talent: ν_1 is intermediate in the sense that the Yesman-*II *equilibrium in Proposition 5.*4 *obtains with the corresponding principal payoff given by* (15). *Then depending on parameter values the dominance can go either way, i.e., the sign of* (15) − (8) *can be positive or negative, as the second panel of Fig. 6 indicates.*

Consider Proposition 7.1. The Yesman-II equilibrium (see Proposition 5.2) will have the unbiased leader virtue signaling and the biased leader is cut-throat, whereas under the Yesman-I equilibrium (Proposition 4.1(C)) the unbiased leader is again virtue signalling and the biased leader is signal respecting. Thus Yesman-II performs worse in period 1 project choice relative to Yesman-I (cut-throat vs. signal respecting implementation by the biased leader) but performs better in period 2 (due to superior screening). The period 2 surplus effect is however small because in case talent wins the signal received by an honest agent 2 is not lost, which is an important driver of period 1 surplus given that ν_2 is intermediate. However, the period 1 effect dominates because, with ν_1 intermediate,

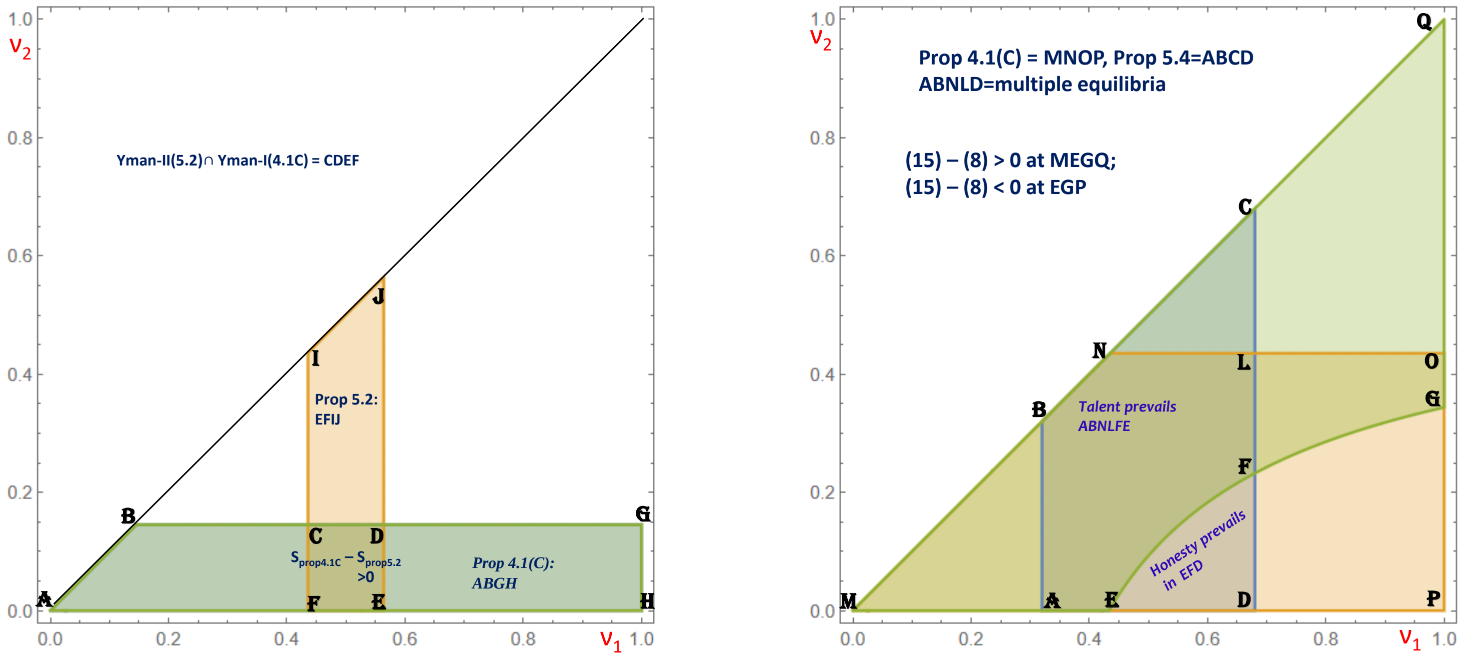


Figure 6: First panel: Surplus comparison (Proposition 7.1) – Proposition 4.1(C) vs. Proposition 5.2, fixing $\xi = 0.6, \lambda = 0.8, \epsilon = 0.05, s = 0.1, V_1 = 1000, V_2 = 4750, \zeta \geq 340$. So in the region CDEF where the two equilibria overlap, honesty is preferred over talent. In numerical computation $S_{prop4.1C} - S_{prop5.2} > 0$ for the entire area ABGH. Second panel: Surplus comparison (Proposition 7.2) – Proposition 5.4 vs. Proposition 4.1(C), fixing $\xi = 0.6, \lambda = 0.8, \epsilon = 0.05, s = 0.1, V_1 = 1100, V_2 = 2500, \zeta \geq 340$.

the fact that a biased leader is signal-respecting in case talent wins, adds significantly to the period 1 surplus differential.

Why can talent win when the Yesman-II equilibrium involves that in Proposition 5.4? Under this equilibrium the period 1 outcome involves both the biased, and the unbiased leader being signal-respecting, which maximizes period 1 surplus. While there is no effective screening since regime-switching is random, the fact that the honesty levels of the two agents are not too divergent ensures that the period 2 screening effect is not too large. This ensures that there exists parameter values such that talent wins.

Remark 2 *What happens if Assumption 3 does not hold, i.e. complete pooling is allowed, in particular for parameter values such that a complete pooling equilibrium exists under the Yesman-*I *scenario, and the Yesman-*II *equilibria involve either the one in Proposition 6, or the one in Proposition 7? In Bag and Roy Chowdhury (2022) we demonstrate that in either case talent wins, even when honesty would have been preferred if complete pooling is ruled out. Thus allowing for the complete pooling equilibrium makes it more likely that talent wins, which is intuitive given that it leads to the least possible surplus compared to any other Yesman-*I *equilibrium that may exist (Lemma 3).*

6.1 Honesty differential small

The choice between talent and honesty becomes more nuanced when the honesty differential is small, with the period 1 leadership choice depending on combinations of (small) talent differential, difference in honesty, relative valuations of period 1 and period 2 projects, and the noise in signal reception. We will show that the principal will make the leadership decision based on the *worth* of the two agents, a composite indicator that depends both on the level of *net talent*, $\tau_i - \frac{1}{2}(1-\epsilon)$, and the degree of unbiasedness, ν_i, $i = 1, 2$. It is sufficient to consider the case where ν_1 and ν_2 are both large.[31] We start with a lemma.

Lemma 6 *The surplus difference* (12) − (6)

$$= (1-\epsilon)\frac{1}{2}\Big\{\nu_2(2\lambda-1) - \nu_1(2\xi-1)\Big\}(s_P V_1) + \frac{1}{2}(\nu_1-\nu_2)\big[(1-\epsilon)X - \frac{1}{2}\big](s_P V_2)$$
$$= \Big[W(\nu_2,\tau_2) - W(\nu_1,\tau_1)\Big](s_P V_1) + \frac{1}{2}(\nu_1-\nu_2)\big[(1-\epsilon)X - \frac{1}{2}\big](s_P V_2),$$

which can be positive or negative. See Fig. 7.

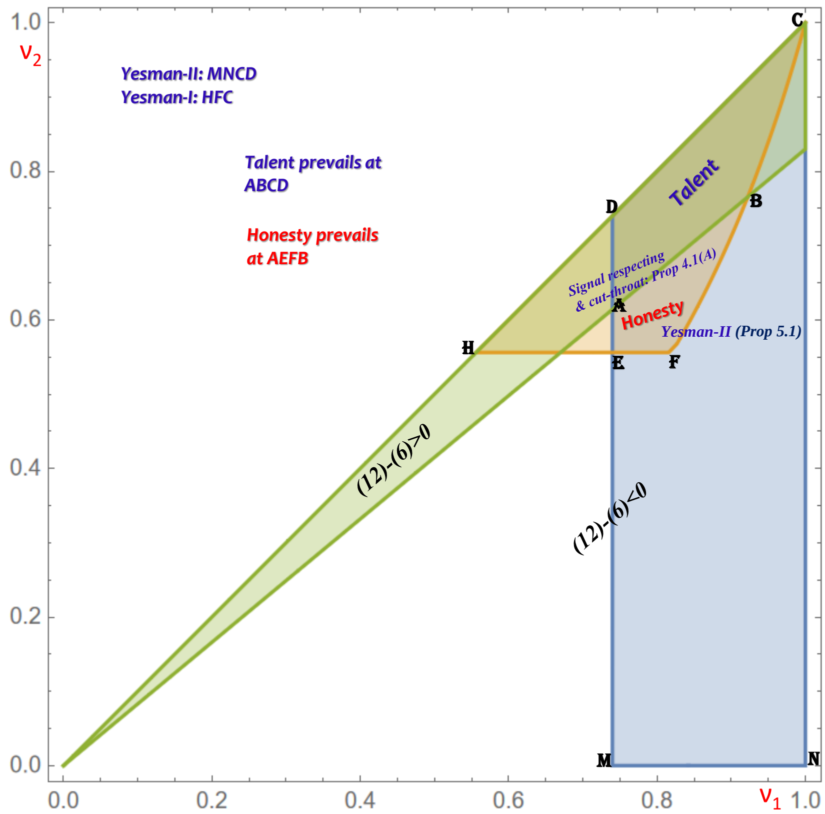


Figure 7: Either talent or honesty can dominate [Proposition 8], fixing $\xi = 0.65, \lambda = 0.75, \epsilon = 0.45, s = 0.1, V_1 = 1000, V_2 = 3870, \zeta \geq 270$.

Proposition 8 ***(*****Small difference in honesty*****)*** *Suppose that ν_1 is large, so that the Yesman-*II *equilibrium in Proposition 5.1 obtains with the resulting surplus (12). Further, let ν_2 be large, i.e. (4) holds, and (5) is violated, and moreover, $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} < \nu_2$, so that Yesman-*I *equilibrium in Proposition 4.1(A) obtains with the surplus given by (6). Then either of talent or honesty could be the decisive factor for leadership in period 1 (see Fig. 7):*

[31] The analysis for the other cases, which yield qualitatively similar results, are available on request.

1. *The principal appoints the talented agent 2 as the period 1 leader, i.e.* (12)−(6)> 0, *whenever the worth of agent 2 exceeds that of agent 1, i.e.* $\nu_2(\lambda - 1/2) - \nu_1(\xi - 1/2) > 0$.

2. *The principal appoints the more honest agent 1 as the period 1 leader, i.e.* (12)−(6)< 0, *whenever the worth of agent 1 exceeds that of agent 2, i.e.* $\nu_2(\lambda - 1/2) - \nu_1(\xi - 1/2) < 0$ *and the talent signal is noisy, formally* $\epsilon > \bar{\epsilon}$, *where* $\bar{\epsilon} < 1$.

The intuition is as follows. Note that the Yesman-I equilibrium in Proposition 4.1(A) and the Yesman-II equilibrium in Proposition 5.1 have identical structures – the unbiased leader signal respecting and the biased leader cut-throat. Hence, period 1 surplus is higher under the Yesman-I equilibrium if agent 1 has a higher worth relative to that of agent 2. Turning to period 2 surplus however, note that the probability that an unbiased agent assumes period 2 leadership under Yesman-I is $(1-\nu_1)\nu_2+\nu_1[1-(1-\epsilon)\{\frac{1}{2}\xi+\frac{1}{2}(1-\xi)\}]+\nu_1(1-\epsilon)\frac{1}{2}\nu_2 = (1-\nu_1)\nu_2+\nu_1[1-\frac{1}{2}(1-\epsilon)(1-\nu_2)]$, whereas the corresponding probability under Yesman-II is $(1-\nu_2)\nu_1+\nu_2[1-(1-\epsilon)\{\frac{1}{2}(1-\lambda)+\frac{1}{2}\lambda\}]+\nu_2(1-\epsilon)\frac{1}{2}\nu_1 = (1-\nu_2)\nu_1+\nu_2[1-\frac{1}{2}(1-\epsilon)(1-\nu_1)]$. So in period 2, $\text{Surplus}_{YM-II}-\text{Surplus}_{YM-I} = (1/2)(1-\epsilon)(\nu_1-\nu_2) > 0$. Thus there is a tradeoff between the two leadership choices. Whenever agent 2 has a higher worth, even period 1 surplus is higher with agent 2 as the leader and thus talent wins. Whereas if agent 1 has a relatively higher worth, then as the signal becomes noisy the screening advantage of the Yesman-II equilibrium gets smaller, vanishing when $\epsilon = 1$. In that case honesty wins.

7 Truth-telling Deputy: Does this necessarily Increase Surplus?

It is clear that Yesman equilibria are inefficient since the deputy suppresses all signals in period 1, so that there is significant loss of information. Would other equilibria where the deputy reveals more information *necessarily* improve the principal's surplus? Somewhat surprisingly we find that the answer is in the negative.

To that end we consider a scenario where the leader is untalented, with talent ξ, and the deputy is talented, with talent λ. We then construct an equilibrium that we call an *informative advice equilibrium* where the deputy truthfully reports all signals, and the leader always follows the deputy's report. Note that this equilibrium is in some sense the polar opposite of Yesman equilibria in that the deputy reveals all information. Even so, we find that the principal's surplus under the corresponding Yesman-I equilibrium may dominate the principal's surplus under the informative advice equilibrium.

We start with a formal definition of this equilibrium in the continuation game starting at T_{12}.

■ *Informative advice equilibrium:*

Period 1. *Stage* T_{12}. The deputy truthfully reports all signals.

Stage T_{13}. The leader implements all non-null reports by the deputy. Further, in case of a null report, if her own signal is null she implements b_1 if she is honest, and a_1 otherwise. On the other hand, if the deputy submits a null report but her own signal is non-null, she follows her signal (i.e., signal respecting irrespective of bias type).

Period 2. *Stage* T_{21}. The *principal's* strategy depends on whether the deputy makes a null recommendation or not. If the deputy makes a null recommendation, then the principal's strategy is to replace period 1 leader if she implements the project favored by her corrupt self, i.e., a_1, and otherwise retain the leader for period 2. If the deputy makes a non-null recommendation, then there is regime change if and only if the leader implements a project that goes against the deputy's recommendation.

Stage T_{22}. The *deputy's* recommendation strategies are same as the one by period 1 deputy.

Stage T_{23}. The *leader's* project choice follows the sincere equilibrium strategies as in Proposition 1.

Beliefs: The posterior beliefs about the deputy's types will be determined using Bayes' rule. There is no out-of-equilibrium report by the deputy in period 1 or period 2. ‖

Define

$$\overline{V}_2(\epsilon) \equiv \frac{\max\{2\xi - 1, \frac{\lambda-\xi}{(1-\xi)\lambda+\xi(1-\lambda)}\}}{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2} - \frac{1}{2}} V_1.$$

Proposition 9 *(***Existence of informative advice equilibrium***)* *There exists $\epsilon'' > 0$ and $\nu_1'' < 1$ such that, for all $\epsilon < \epsilon''$, $\nu_1 > \nu_1''$, $2\overline{V_2}(\epsilon) > V_2$, and $sV_2 > \max\{[(1-\lambda)(1-\xi)+\xi\lambda]\zeta, [(1-\lambda)\xi+(1-\xi)\lambda]\zeta\}$, there exists an interval $(\underline{\nu}, \overline{\nu})$, where $\underline{\nu} < \frac{1}{2} < \overline{\nu}$, such that for all $\nu_2 \in (\underline{\nu}, \overline{\nu})$ an informative advice equilibrium exists.*

In this equilibrium the principal's surplus is given by:

$$X(s_P V_1) + \left[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\frac{\xi}{4}\right]\left(\nu_1 X + (1-\nu_1)\frac{1}{2}\right)(s_P V_2) \\ + \left[\frac{\epsilon}{2} + \frac{\epsilon^2}{2}\right]\left(\nu_2 X + (1-\nu_2)\frac{1}{2}\right)(s_P V_2) + \epsilon^2 X(s_P V_2). \tag{17}$$

Proposition 10 *(***Beneficial backstabbing***)* *Suppose an informative advice equilibrium exists.*

1. *If $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} < \nu_2$, (4) holds, and (5) is violated, so that the Yesman-I equilibrium in Proposition 4.1(A) obtains, and the principal's surplus is given by (6). Depending on parameter values the principal's surplus in the Yesman-I equilibrium may exceed that in the informative advice equilibrium (i.e., (17)), as well as the other way around. See Fig. 12.*

2. *If $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} \geq \nu_2$, Yesman-I equilibrium in Proposition 4.2 obtains, and the principal's surplus is given by (9). Depending on parameter values the principal's surplus in the Yesman-I equilibrium may exceed that in the informative advice equilibrium, as well as the other way around. See Fig. 13.*

The intuition for the ambiguous ranking in Proposition 10.1 can be understood by considering the case of $\epsilon = 0$.[32] Under the informative advice equilibrium period 1 implementation is informationally efficient, but there is a screening inefficiency in that a biased leader continues to be the leader in period 2 with probability $1-\nu_1$ which pulls down period 2 payoff.[33] In contrast, under the Yesman-I equilibrium in Proposition 4.1(A), the period 1 outcome is informationally inefficient, but the same inefficiency improves period 2 screening, demonstrating the negative relationship between period 1 and

[32]For $\epsilon > 0$, similar countervailing forces will prevail.

[33]Under informative advice period 2 leader is biased with probability $(1-\epsilon)(1-\nu_1) + \epsilon^2(1-\nu_1)(1-\nu_2) + \epsilon(1-\epsilon)\frac{1}{2}(2-\nu_1-\nu_2)$. By setting $\epsilon = 0$, this probability equals $1-\nu_1$.

period 2 efficiency that we flagged earlier. The inefficiency in period 1 under Yesman-I happens both because the deputy makes an uninformative report, and also because the biased leader is cut-throat, both potentially leading to wrong project implementation.[34] Thus while the Yesman-I equilibrium has a lower surplus in period 1, it has a higher period 2 surplus relative to the informative advice equilibrium. Hence, the Yesman-I equilibrium is more likely to have a higher surplus whenever V_2 is relatively large. Overall, the resulting ambiguity of ranking is shown in an example in Fig. 12 for $\epsilon > 0$.

Next consider part 2. Here we make a distinction between $\epsilon = 0$ and $\epsilon > 0$. The performance of the informative advice equilibrium is already explained above: efficiency of period 1 implementation,[35] but inefficient implementation in period 2 with probability $1 - \nu_1$ when $\epsilon = 0$. Clearly, under the Yesman-I equilibrium in Proposition 4.2, period 1 implementation is less efficient, whereas in period 2, the probability of the leader being biased is again $1 - \nu_1$. Thus the informative advice equilibrium necessarily yields a higher surplus when $\epsilon = 0$.

We next demonstrate that, for ϵ positive, a tradeoff again emerges, with the Yesman-I equilibrium in Proposition 4.2 having a more efficient screening mechanism for ϵ positive, but not too large. Consider the informative advice equilibrium, and suppose the deputy submits a null report. Then the current leader is replaced if either she is biased, observes no signal and implements a_1, or she may be replaced if she observes a signal. The key intuition comes from the second possibility, so let us decompose the probability that the period 2 leader is biased in this case:

$$\epsilon(1-\epsilon)[\frac{1-\nu_1}{2} + \frac{\nu_1}{2}(1-\nu_2) + \frac{(1-\nu_1)(1-\nu_2)}{2}] = \epsilon(1-\epsilon)\frac{2-\nu_1-\nu_2}{2}, \tag{18}$$

where $\epsilon(1-\epsilon)$ is the probability that the leader has a signal while the deputy has none, $\frac{1-\nu_1}{2}$ is the probability that the biased leader has signal β, implements b_1 and retains leadership, $\frac{\nu_1}{2}(1-\nu_2)$ is the probability that an unbiased leader has signal α, implements a_1, and is replaced by a biased agent 2, and finally $\frac{(1-\nu_1)(1-\nu_2)}{2}$ is the probability that the biased leader observes α, implements a_1 and is replaced by a biased agent 2. Note that $\frac{2-\nu_1-\nu_2}{2}$ exceeds $1-\nu_1$, which is the period 2 bias probability under the Yesman-I equilibrium. Thus this term would tend to dominate whenever $\epsilon(1-\epsilon)$ is large, which happens for ϵ positive, but not too large. This is the reason why we have the possibility that the Yesman-I equilibrium may outperform informative advice equilibrium. Again this is more likely if V_2 is larger. This possibility is illustrated in Fig. 13.

Note that here we only provide a partial analysis of the informative advice equilibria. For one, we do not analyze the informative advice equilibrium where the talented agent is the leader, etc. This is because of two reasons. First, given our focus on the Yesman equilibria, a full analysis of the informative advice equilibrium is beyond the scope of this paper. Second, and relatedly, our objective in this section is simply to make the point that various strategic considerations ensure that even full

[34] The relatively superior performance of Yesman-I in period 2 can be tracked as follows: the leader (in period 2) is biased with probability $\frac{1}{2}\nu_1(1-\nu_2)+(1-\nu_1)(1-\nu_2)$ (using strategies in Lemma 1 and Proposition 4.1(A)), which is less than $(1-\nu_1)$, the corresponding probability under informative advice; the key advantage of Yesman-I is that a biased period 1 leader with a non-null signal is always replaced in period 2 (due to cut-throat strategy), which is not necessarily the case under informative advice (signal respecting strategy). Note that $(1-\nu_1) > \frac{1}{2}\nu_1(1-\nu_2) + (1-\nu_1)(1-\nu_2)$, i.e. $\nu_1 < \frac{2\nu_2}{1+\nu_2}$, follows from the given condition for Proposition 4.1(A) by setting $\epsilon = 0$.

[35] It is easy to see from the equilibrium strategies that period 1 implementation continues to be ex-ante efficient even for $\epsilon > 0$.

information revelation by the deputy need not dominate Yesman equilibria in terms of the principal's surplus.

The preceding results have implications for a broader debate in organizations and politics – the efficiency implications of *backstabbing* by a second-in-command.[36] An important form of backstabbing, especially in the workplace, is information suppression, as manifested in the Yesman strategies that we study in this paper.[37] In general backstabbing has a negative connotation. For example, in team and promotion tournament problems Konrad (2000) and Chen (2003) discuss the issue of sabotage, a form of backstabbing, finding that it plays a negative role. One important contribution of this paper is to unearth a *positive* role for backstabbing – as an identifier of corruption and thus in culling out corrupt leaders. In fact, our analysis suggests that given the choice the principal may sometimes prefer that the deputy behaves in a backstabbing manner by withholding information. Interestingly the role of strategic concealment of information has been highlighted by various commentators, in the organizational context by Detert and Edmond (2012)[38] and in the context of career concerned politicians by Diermeier et al. (2005).

8 Conclusion

We advance the literature on project implementation through delegation by drawing upon the broad insights from the literature on reputation games and cheap talk advice. What has been missing so far is formalization of one of the key attributes of a leader, *the honesty* (or integrity), in performing the role expected of the leader. In any job where the leader has an absolute authority vested in her, in choosing projects that make a big difference to the organization's value, honesty is as important as the leader's talent. Talent without honesty and vice versa can be the undoing of big initiatives. But when honesty (or bias) is added as a second attribute, to the literature's well-studied role of talent, the interaction between these two attributes in decisions on delegation of authority, advice and project implementation become quite complex. This paper tackles a new set of issues with novel intricacies using a two-period dynamic model of leader-deputy interactions where the leader can lose her position to the deputy. The work is applicable to organizations and politics where decisions are made by multiple experts with a hierarchy of authority and advice.

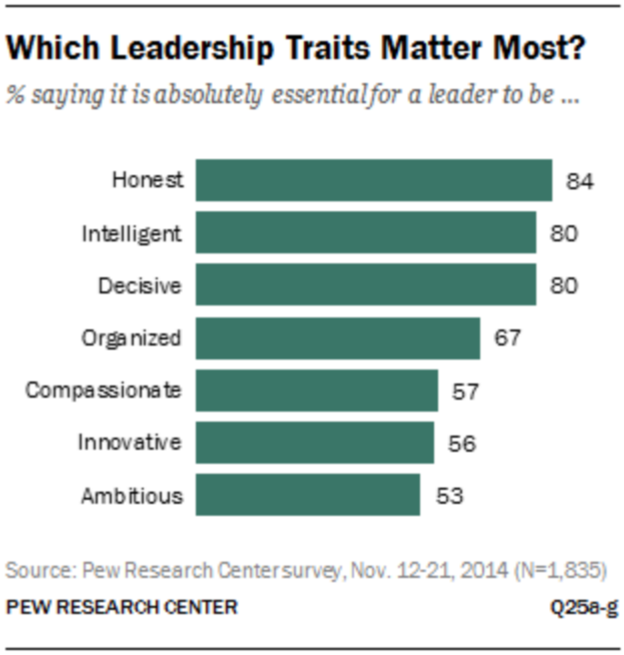
The model is kept simple in some of the aspects. For example, which project a corrupt expert favors is assumed to be common knowledge. In many applications such an assumption is reasonable because of the experts' past interactions with the same parties who are currently involved, or due to intense lobbying by particular projects' developers. The projects are assumed to be exogenous, and the only beneficiary from corruption is the leader who has the ultimate authority on project selections. The replaced leader will have the same stakes in future choices and does not engage in spiteful acts, which can be a good description for organizations and political applications at least in the medium term. At the minimum, the experts will have future careers and thus performing well in their current positions is a valid enough motivation. Finally, in keeping the analysis tractable we

[36]The same *Yesman* behavior that can be labelled as being conformist can also be seen as being devious.

[37]In their study of backstabbing in workplace, Malone and Hayes (2012) identify "Withheld or Concealed Information' as an important category of such backstabbing.

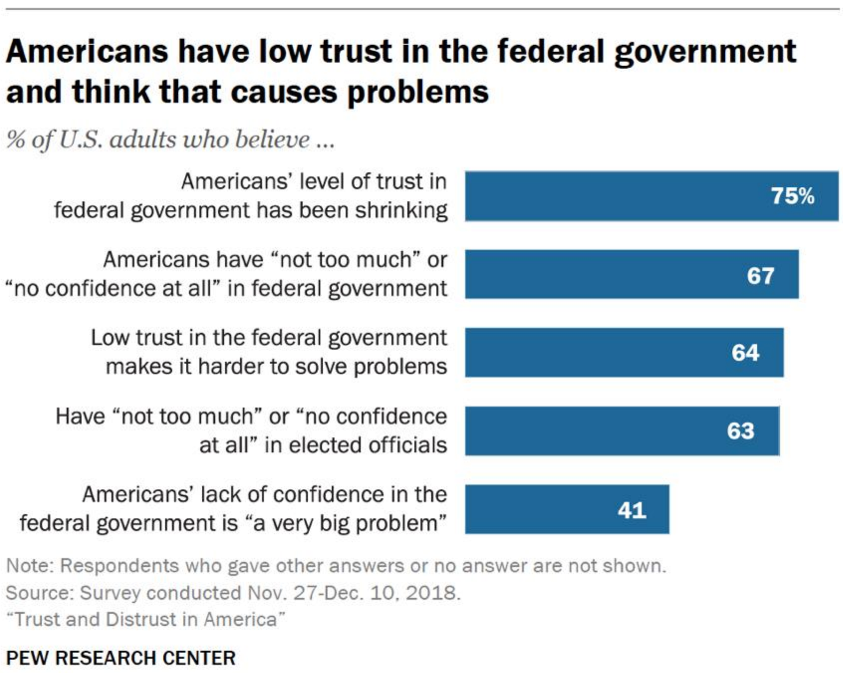
[38]To quote, "When in doubt, keep your mouth shut."

assumed uncertainty only in the experts' biases. This allowed us to develop the key tradeoffs between honesty and talent, a central focus of this paper.[39]



https://www.pewresearch.org/social-trends/2015/01/14/chapter-2-what-makes-a-good-leader-and-does-gender-matter/

Figure 8: Leadership traits



How Americans see problems of trust; source https://www.pewresearch.org/politics/2019/07/22/how-americans-see-problems-of-trust/

Figure 9: "Confidence in leaders and major institutions is mixed"

[39]In an earlier work (Bag and Roy Chowdhury, 2021), we considered both talent and bias to be uncertain. There while some of the forces driving the principal's leadership choices are similar in nature, a clean characterization proved difficult due to the more complex strategic interactions.

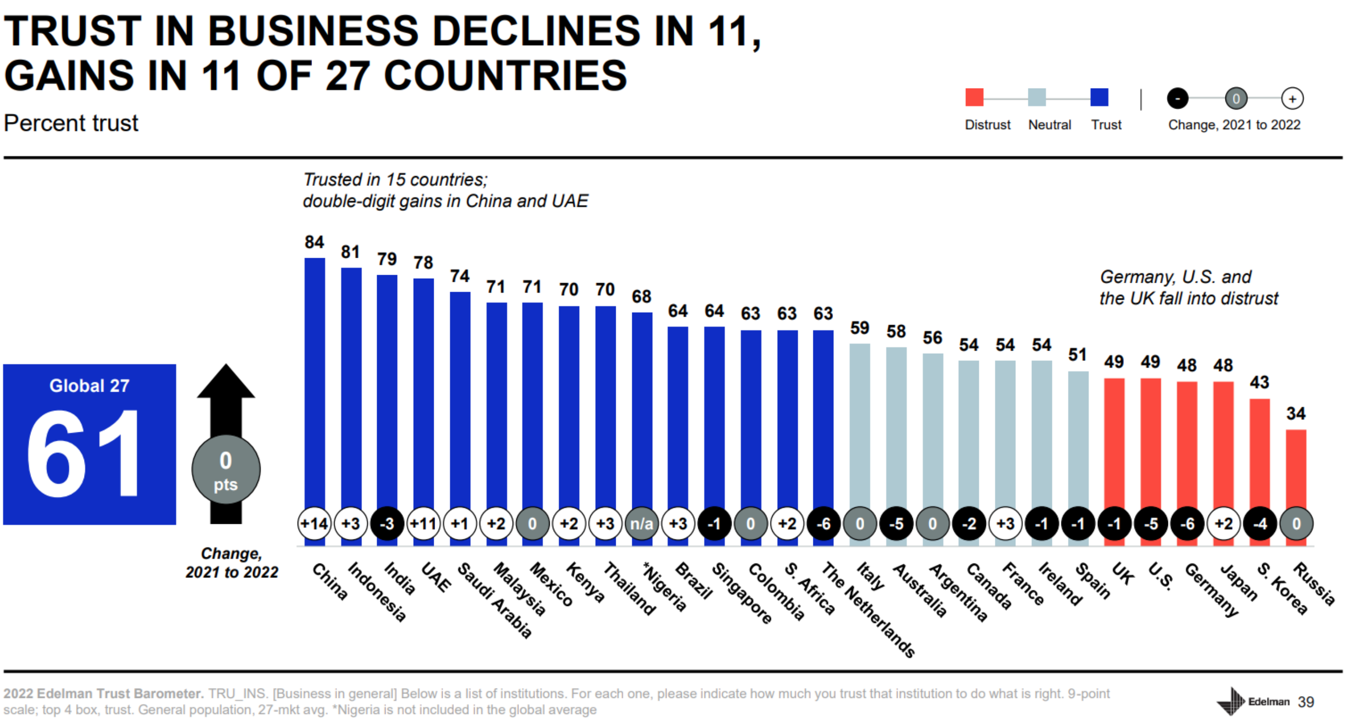


Figure 10: Edelman barometer: Falling trust in business

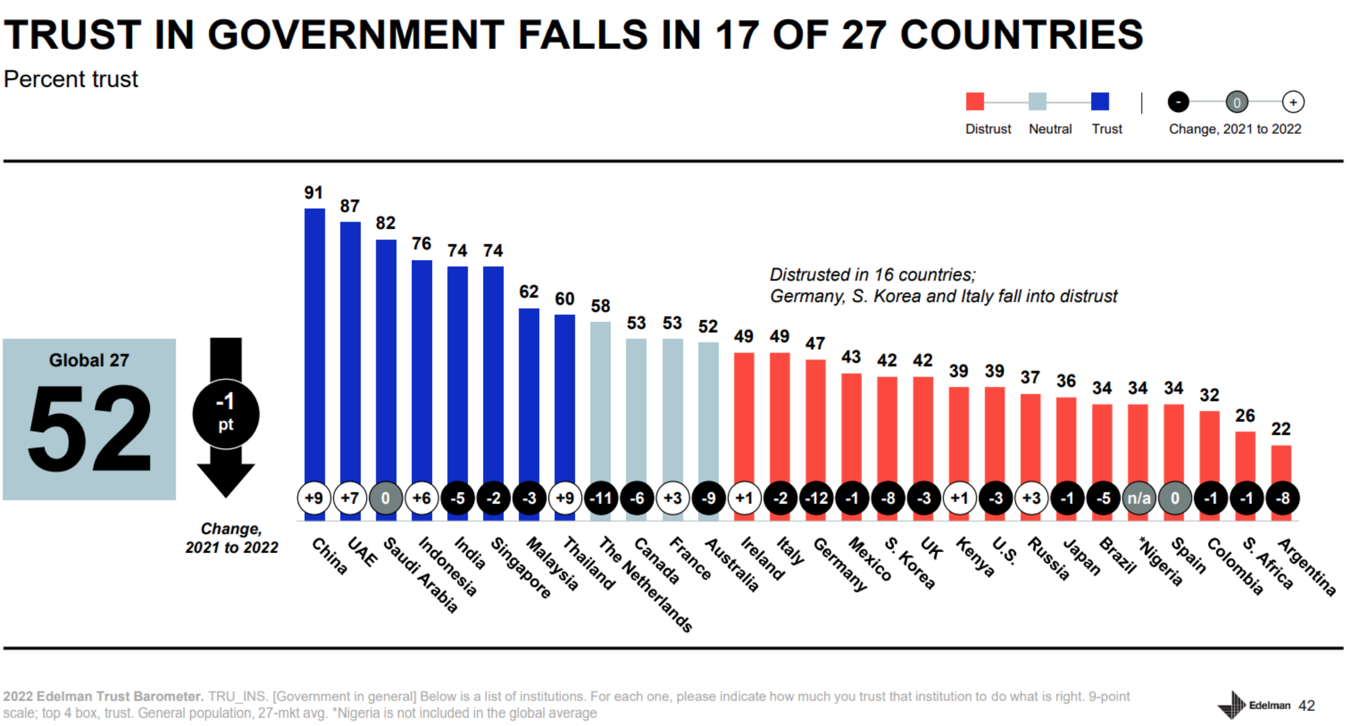


Figure 11: Edelman barometer: Falling trust in governments

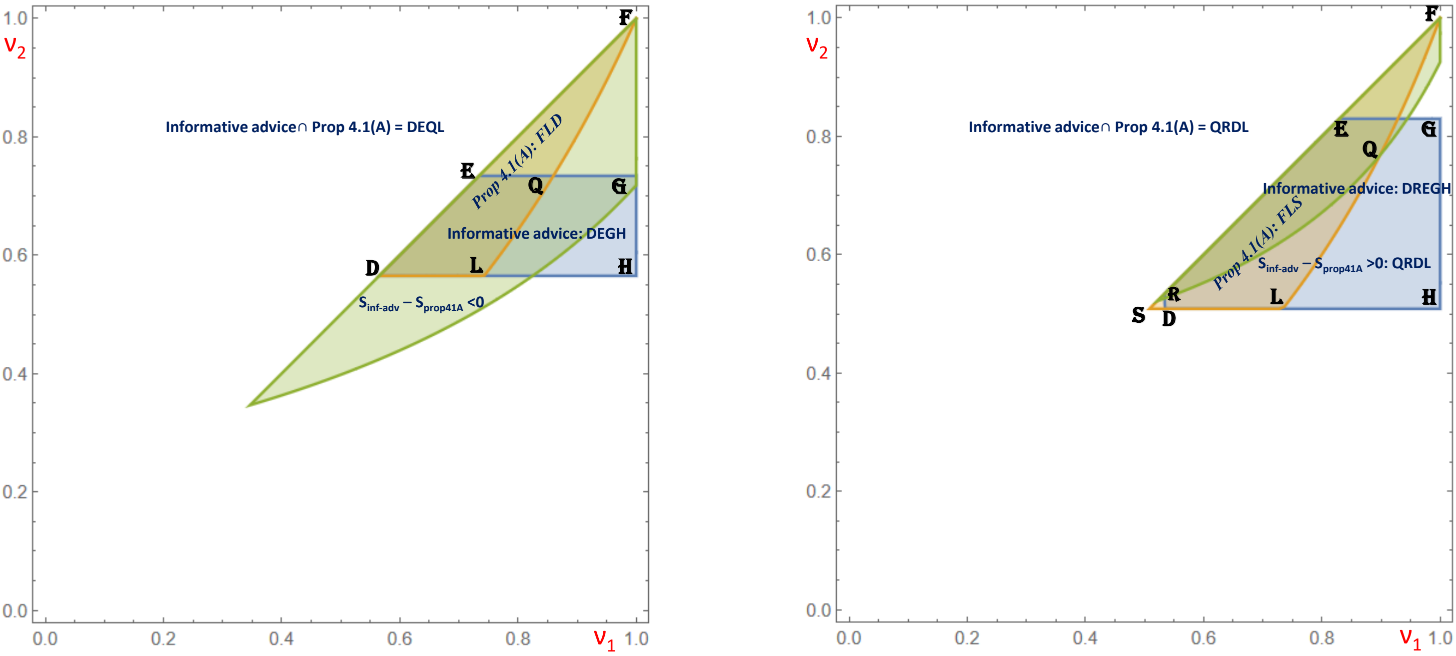


Figure 12: First panel: Surplus difference<0, Informative advice, Proposition 4.1(A) equilibria, fixing $\xi = 0.65, \lambda = 0.85, \epsilon = 0.1, s = 0.1, V_1 = 2000, V_2 = 4200, \zeta = 350$. Second panel: Surplus difference>0, Informative advice, Proposition 4.1(A) equilibria, fixing $\xi = 0.65, \lambda = 0.85, \epsilon = 0.25, s = 0.1, V_1 = 2000, V_2 = 4200, \zeta = 350$.

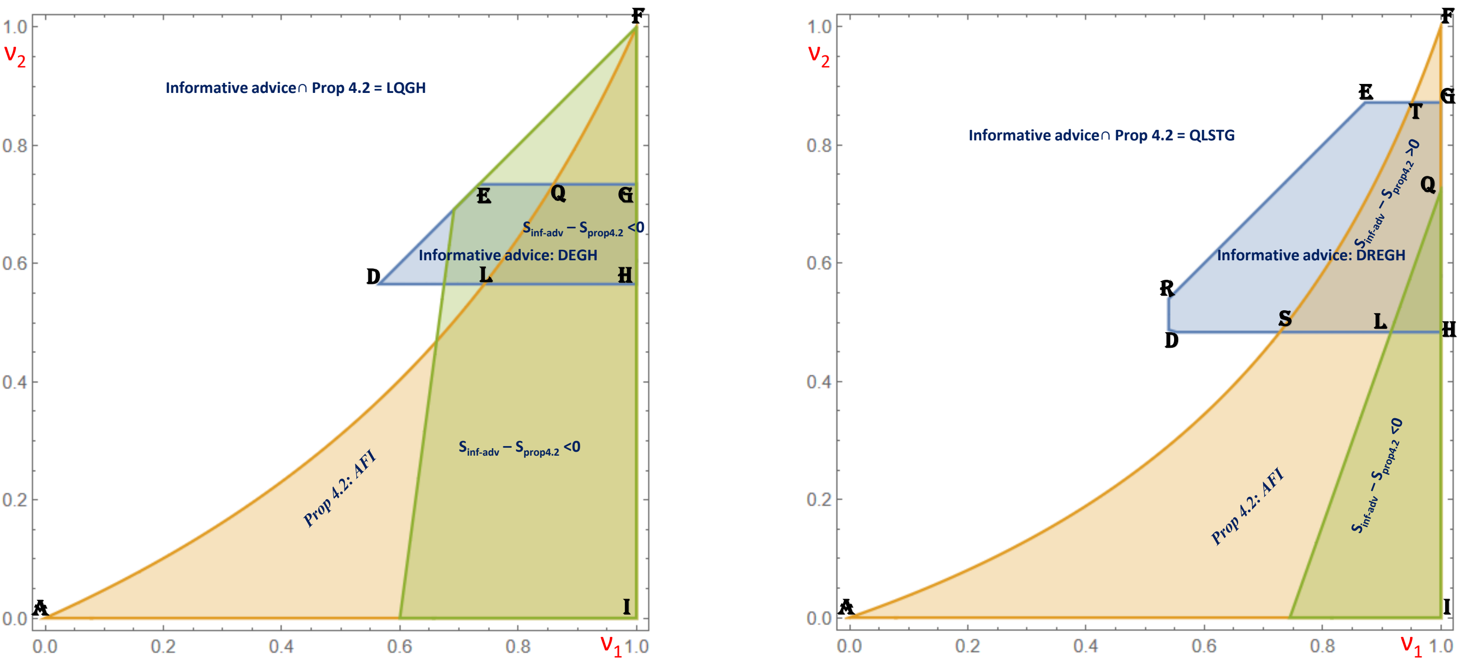


Figure 13: First panel: Surplus difference<0, Informative advice, Proposition 4.2 equilibria, fixing $\xi = 0.65, \lambda = 0.85, \epsilon = 0.1, s = 0.1, V_1 = 2000, V_2 = 4200, \zeta = 350$. Surplus difference$>0$, Informative advice, Proposition 4.2 equilibria, fixing $\xi = 0.65, \lambda = 0.85, \epsilon = 0.3, s = 0.1, V_1 = 2000, V_2 = 4200, \zeta = 350$.

A Appendix

Proof of Proposition 1. First consider the deputy. In period 2 the deputy's bias has no relevance for her payoff. If the leader is biased, the deputy's report is irrelevant, so the deputy tells the truth. An unbiased leader would take the deputy's information into account, so the deputy obtaining a fixed share of the expected surplus would report truthfully.

Next consider the leader. Suppose that expert 1 is the period 2 leader and biased. It suffices to examine the case where both the leader and the deputy observe β, i.e., the signal going against the leader's pet project. The leader's expected payoff from not following this common signal and instead implementing her pet project a_2 is $\zeta + \frac{(1-\lambda)(1-\xi)}{\lambda\xi+(1-\lambda)(1-\xi)}(sV_2)$, while that from implementing b_2 according to the common signal is $\frac{\lambda\xi}{\lambda\xi+(1-\lambda)(1-\xi)}(sV_2)$. Comparing these two payoffs and applying Assumption 2(a), the claim follows. (Apply the same logic for expert 2 as the leader.) If the leader is unbiased, given that the deputy reports truthfully, the leader takes the deputy's as well as her own information into account, which ensures that the project selection is efficient and so her own payoff is maximized. **Q.E.D.**

Proof of Proposition 2. **1.** From Proposition 1, in period 2 a biased leader chooses project a_2 (b_2) if she is expert 1 (resp. expert 2). Call the expected (total) surplus from the project (ignoring gains from bias, as well as B) with a biased expert i as the leader to be $S_{i,2}(\neg N)$. For $q_t = \frac{1}{2}$, $S_{1,2}(\neg N) = q_2 V_2 = \frac{V_2}{2}$. In case expert 2 is the leader she always chooses project b_2. Thus $S_{2,2}(\neg N) = (1-q_2)V_2 = \frac{V_2}{2}$. Consequently, $S_{1,2}(\neg N) = S_{2,2}(\neg N) = S_2(\neg N)$. In case the leader is unbiased, the project selection is efficient so the expected surplus is again independent of the agent's identity. In that case we call the expected surplus $S_2(N)$. Thus the principal's expected payoff when expert i is the leader is:

$$s_P[(1-\mu_i(P))S_2(\neg N) + \mu_i(P)S_2(N)] = s_P[(S_2(\neg N) + \mu_i(P)\{S_2(N) - S_2(\neg N)\}]. \tag{A.1}$$

With informative signals, $S_2(N) > S_2(\neg N)$. So from (A.1), principal's payoff is increasing in the degree of unbiasedness $\mu_i(P)$ and he selects the expert with a higher $\mu_i(P)$.

2. Given that the period 2 equilibrium is a sincere one, the principal is only interested in comparing the expected bias level of the two agents. We argue that project choice is a sufficient statistic to that end. Define three events A, B, and C, such that A is the event that the agent, either 1 or 2, is biased, B denotes project choice, i.e. whether a_1, or b_1 is chosen, and C denotes the outcome of the selected project, i.e. whether it is a success, or not. We need to show that:

$$P(A|B \cap C) = P(A|B).$$

It is clear that (a) $P(A \cap B|C) = P(A \cap B)$, and (b) $P(B|C) = P(B)$. Next note that

$$P(A|B \cap C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} = \frac{P(A \cap B|C)P(C)}{P(B|C)P(C)} = \frac{P(A \cap B)}{P(B)} = P(A|B),$$

where the third equality uses the fact that $P(A \cap B|C) = P(A \cap B)$, and $P(B|C) = P(B)$, and that $P(C) \neq 0$ by assumption. **Q.E.D.**

Proof of Lemma 1. **(i)** Consider an (N, β) leader. Given that her signal is β, a deviation to a_1 lowers her period 1 payoff, and moreover triggers a regime switch which lowers her period 2 payoff as well (because the period 2 leader could be biased). Thus it is optimal to implement b_1.

Next consider an $(N, \emptyset)$ leader. She prefers to implement b_1, since project choice does not affect her period 1 payoff, but implementing b_1 generates a higher period 2 payoff since there is no regime switch.

(ii) Note that (4) is equivalent to

$$\xi(sV_1) + \big[\nu_2 X + (1-\nu_2)\frac{1}{2}\big](sV_2) \geq (1-\xi)(sV_1) + X(sV_2), \tag{A.2}$$

where X is as defined in (3), the LHS is her payoff from implementing a_1, and the RHS is her payoff from implementing b_1.

(iii) Consider a biased leader having a signal α. She prefers to implement a_1 if and only if

$$\zeta + \xi(sV_1) + \big[\nu_2 X + (1-\nu_2)\frac{1}{2}\big](sV_2) \geq (1-\xi)(sV_1) + \big\{\zeta + \frac{1}{2}(sV_2)\big\}, \tag{A.3}$$

where the LHS is her payoff from implementing a_1, and the RHS is her payoff from implementing b_1. The inequality can be re-written as $(2\xi - 1)sV_1 + \nu_2[X - \frac{1}{2}](sV_2) \geq 0$, which is always satisfied since $\xi > \frac{1}{2}$, and $X - \frac{1}{2} > 0$.

Next consider a biased leader having a signal $\emptyset$. She prefers to implement a_1 if and only if

$$\zeta + \frac{1}{2}(sV_1) + \big[\nu_2 X + (1-\nu_2)\frac{1}{2}\big](sV_2) \geq \frac{1}{2}(sV_1) + \big\{\zeta + \frac{1}{2}(sV_2)\big\}, \tag{A.4}$$

where the LHS is her payoff from implementing a_1, and the RHS is her payoff from implementing b_1. The inequality can be rewritten as $\nu_2[X - \frac{1}{2}](sV_2) \geq 0$, which again is always satisfied.

(iv) Note that (5) is equivalent to

$$\zeta + \xi(sV_1) + \frac{1}{2}(sV_2) \geq \zeta + (1-\xi)(sV_1) + \big[\nu_2 X + (1-\nu_2)\frac{1}{2}\big](sV_2), \tag{A.5}$$

where the LHS is her payoff from implementing b_1, and the RHS is her payoff from a_1. **Q.E.D.**

Proof of Lemma 2. **(i)** Note that given the principal's strategy, project choice in period 1 does not affect period 2 payoffs. Thus an unbiased leader always goes with her signal since it maximizes her period 1 payoff. Whereas, given that $\zeta > (2\xi - 1)sV_1$, the biased leader implements her favored project a_1 even if the signal is β, and consequently if the signal is either α, or $\emptyset$. Note however, that on observing b_1, the principal infers that the leader is unbiased (since a biased leader never implements a_1) and does not replace her in period 2. This is a contradiction.

(ii) We can mimic the argument in part (1) of this lemma to show that an unbiased leader always goes with her signal, and the biased leader implements her favored project a_1 even if the signal is β, and consequently if the signal is either α, or $\emptyset$. Moreover, given the tie-breaking rule, the unbiased leader implements b_1 on obtaining a null signal. Note that on observing a_1, the principal's posterior regarding the honesty of the period 1 leader decreases to $\frac{\nu_1(1-\epsilon)}{2-\nu_1-\nu_1\epsilon}$. This follows because a biased leader always implements a_1, whereas an unbiased leader implements a_1 if and only if her signal is β. Given that $\frac{\nu_1(1-\epsilon)}{2-\nu_1-\nu_1\epsilon} < \nu_2$, the principal necessarily replaces her on observing a_1. This is a contradiction.

(iii) Suppose the principal's strategy involves regime switch if and only if b_1 is implemented. Given the principal's strategy the biased leader always chooses a_1. Next consider an unbiased leader. She implements a_1 on observing α (as that is more profitable), and also on observing $\emptyset$ since doing so prevents a potentially biased agent 2 from obtaining period 2 leadership. Whereas on observing β, she implements a_1 if and only if $(2\xi - 1)sV_1 < (1 - \nu_2)[X - \frac{1}{2}]sV_2$.

Given that $(2\xi-1)sV_1 \geq (1-\nu_2)[X-\frac{1}{2}]sV_2$, the unbiased leader always implements b_1 on observing β. However, in that case b_1 signals that the leader is unbiased and should not be replaced. This is a contradiction. **Q.E.D.**

Proof of Proposition 4. **1(A)** We first write down the candidate equilibrium for this case and then prove that this is indeed an equilibrium:

(i) in period 1, all deputy types send a null report, and the leader adopts the following strategy:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \emptyset\}) \to a_1 \\ (N, \beta) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to a_1 \end{cases}$$

(ii) in period 1, the principal believes that the deputy must be biased with probability 1 whenever the deputy submits a non-null report, but elsewhere the beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal changes period 1 leader if and only if she implemented a_1, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

That the leader and the deputy's period 1 strategies are optimal follows from Lemma 1, and the fact that (4) holds and (5) is violated. Thus it is sufficient to consider the principal's strategy in period 2. From the leader's strategy note that b_1 is implemented only by an unbiased leader. Thus upon observing b_1, the principal concludes that the leader is honest, and does not replace her in period 2. Next consider the case where a_1 is implemented. From the leader's strategy, on observing a_1, the probability that the leader is unbiased decreases to $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)}$. Given that $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} < \nu_2$, it is optimal to replace the leader (Proposition 2, part 1).

1(B) Consider the following candidate equilibrium:

(i) in period 1, all deputy types send a null report, and the leader adopts the following strategy:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \beta, \emptyset\}) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to a_1 \end{cases}$$

(ii) in period 1, the principal believes that the deputy must be biased with probability 1 whenever the deputy submits a non-null report, but elsewhere the beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal changes period 1 leader if and only if she implemented a_1, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

That the leader and the deputy's period 1 strategies are optimal follows from Lemma 1, and the fact that (4) and (5) are both violated. Thus it is sufficient to consider the principal's strategy in period 2. From the leader's strategy note that b_1 is implemented only by an unbiased leader, whereas a_1 is implemented only by a biased leader. Thus upon observing a_1, the principal concludes that the leader is unbiased honest, and does not replace her in period 2, whereas upon observing b_1 the principal concludes that the leader is biased and replaces her.

1(C) Consider the following candidate equilibrium:

(i) in period 1, all deputy types send a null report, and the leader adopts the following strategy:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \beta, \emptyset\}) \to b_1 \\ (\neg N, \beta) \to b_1 \\ (\neg N, \{\alpha, \emptyset\}) \to a_1 \end{cases}$$

(ii) in period 1, the principal believes that the deputy must be biased with probability 1 whenever the deputy submits a non-null report, but elsewhere the beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal changes period 1 leader if and only if she implemented a_1, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

That the leader and the deputy's period 1 strategies are optimal follows from Lemma 1, and the fact that (4) is violated, and (5) holds. Thus it is sufficient to consider the principal's strategy in period 2. From the leader's strategy note that a_1 is implemented only by a biased leader. Thus upon observing a_1, the principal concludes that the leader is biased, and replaces her in period 2. From the leader's strategy, upon observing b_1 the principal's belief regarding the leader's honesty level increases. However, given that $\nu_1 > \nu_2$ to begin with, she is not replaced.

1(D) Finally suppose that (4) and (5) both hold. Hence given Lemma 1, both the unbiased and the biased leader will be signal respecting. Then the leader will not be replaced irrespective of period 1 project implementation. This, however, is a contradiction since the biased leader will then implement a_1 even upon observing β.

2. Consider the following candidate pure strategy PBE where:

(i) in period 1, all deputy types send a null report, and the leader adopts the following strategy:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \emptyset\}) \to a_1 \\ (N, \beta) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to a_1 \end{cases}$$

(ii) in period 1, the principal believes that the deputy must be biased with probability 1 whenever the deputy submits a non-null report, but elsewhere the beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal never replaces the current leader, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

First consider the principal's strategy. On observing b_1 the principal infers that the leader is honest, and thus she is not replaced. Whereas on observing a_1, the principal's posterior regarding the honesty of the period 1 leader decreases to $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)}$. Moreover, the current leader is not replaced if and only if $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} > \nu_1$.

Next given the principal's strategy, it is sufficient to consider period 1 payoffs for the leader. An unbiased leader has a larger payoff if she is signal respecting, and in case of a null signal her expected payoff from both projects are identical. Whereas a biased leader always opts of a_1 even if the signal is β since $\zeta > (2\lambda - 1)sV_1$, and consequently for any other signal.

3. Consider the following candidate pure strategy PBE where:

(i) in period 1, all deputy types send a null report, and the leader adopts the following strategy:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \beta, \emptyset\}) \to a_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to a_1 \end{cases}$$

(ii) in period 1, the principal believes that the deputy must be biased with probability 1 whenever the deputy submits a non-null report, but elsewhere the beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal never replaces the current leader, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

Given the leader strategies, following a_1 the principal believes that the leader is unbiased with probability ν_1, and does not replace her as $\nu_1 > \nu_2$. On observing b_1, all beliefs are permissible since b_1 is off-the-equilibrium. It is clear that given the principal's strategy, and that $(2\xi - 1)sV_1 < (1-\nu_2)[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2 \cdot \frac{1}{2} - \frac{1}{2}]sV_2$, even the unbiased leader with signal β implements a_1.

Finally, an unbiased leader implements a_1 upon observing β if and only if (4) is violated.

4. Consider various possible strategies by the principal.

(i) Suppose the principal's strategy is that there is regime switch if and only if a_1 is implemented. Given the regime-switching rule, from Proposition 4, there are three possible equilibria described in parts $1(A)$, $1(B)$ and $1(C)$. For these parameter values, given Lemma 1, the leader and the deputy's strategies are unique, so that any other equilibria must differ only in the principal's off-the-equilibrium beliefs. Finally, for the parameter space not covered by these three cases, no equilibria exist as shown in part $1(D)$.

(ii) Suppose the principal's strategy is that there is never any regime switch. Given the regime-switching rule, from Proposition 4, there is one possible equilibria described in part 2. For these parameter values, the leader and the deputy's strategies are unique, so that any other equilibria must differ only in the principal's off-the-equilibrium beliefs.

(iii) Suppose the principal's strategy is that there is regime switch if and only if a_1 is implemented. Given the regime-switching rule, from Proposition 4, there are three possible equilibria described in parts $1(A)$, $1(B)$ and $1(C)$. For these parameter values, given Lemma 1, the leader and the deputy's strategies are unique, so that any other equilibria must differ only in the principal's off-the-equilibrium beliefs.

(iv) Finally note that part 1 and part 2 cover the whole parameter space. **Q.E.D.**

Proof of Lemma 4. **(i)** Note that under either candidate strategy of the principal, (a) or (b), project choice in period 1 does not affect period 2 payoffs. Thus an unbiased leader always goes with her signal since it maximizes her period 1 payoff, and, in case of a null signal, opts for a_1 (this follows from her tie-breaking rule). Whereas, given that $\zeta > (2\lambda - 1)sV_1$, the biased leader implements her favored project b_1 even if the signal is α, and consequently if the signal is either β, or $\emptyset$.

Suppose the principal's strategy involves (b), that is there is necessarily a regime switch. Note however that on observing a_1, the principal infers that the leader is unbiased (since a biased leader never implements a_1) and does not replace her in period 2. This is a contradiction.

Next suppose the principal's strategy involves (a), that is there is never a regime switch. Note however, that on observing b_1, the principal's posterior regarding the honesty of the period 1 leader decreases. This follows because a biased leader always implements b_1, whereas an unbiased leader implements b_1 if and only if her signal is β. Given that $\nu_2 < \nu_1$ to begin with, the principal necessarily replaces her. This is a contradiction.

(ii) Given the principal's strategy the biased leader always chooses b_1, so that on observing b_1 the principal's posterior on the period 1 leader being unbiased cannot increase. Given that $\nu_1 > \nu_2$, it is therefore optimal to replace her from leadership. This however is a contradiction, showing that the principal's strategy is not optimal. **Q.E.D.**

Proof of Lemma 5. **(i)** Consider an (N, α) leader. Given that her signal is α, a deviation to b_1 lowers her period 1 payoff, and moreover triggers a regime switch which lowers her period 2 payoff as well (because the period 2 leader could be biased). Thus it is optimal to implement a_1.

Next consider an $(N, \emptyset)$ leader. She prefers to implement a_1, since project choice does not affect her period 1 payoff, but implementing a_1 generates a higher period 2 payoff since there is no regime switch.

(ii) Note that (10) is equivalent to

$$\lambda(sV_1) + \left[\nu_1 X + (1-\nu_1)\frac{1}{2}\right](sV_2) \geq (1-\lambda)(sV_1) + X(sV_2), \tag{A.6}$$

where the LHS is her payoff from implementing b_1, and the RHS is her payoff from implementing a_1.

(iii) Consider a biased leader having a signal β. She prefers to implement b_1 if and only if

$$\zeta + \lambda(sV_1) + \left[\nu_1 X + (1-\nu_1)\frac{1}{2}\right](sV_2) \geq (1-\lambda)(sV_1) + \left\{\zeta + \frac{1}{2}(sV_2)\right\}, \tag{A.7}$$

where the LHS is her payoff from implementing b_1, and the RHS is her payoff from implementing a_1. The preceding inequality can be written as $\nu_1[X - \frac{1}{2}]sV_2 + (2\lambda - 1)sV_1 \geq 0$, which is always satisfied since $\lambda > \frac{1}{2}$, and $X - \frac{1}{2} > 0$.

Next consider a biased leader having a signal $\emptyset$. She prefers to implement b_1 if and only if

$$\zeta + \frac{1}{2}(sV_1) + \left[\nu_1 X + (1-\nu_1)\frac{1}{2}\right](sV_2) \geq \frac{1}{2}(sV_1) + \left\{\zeta + \frac{1}{2}(sV_2)\right\}, \tag{A.8}$$

where the LHS is her payoff from implementing b_1, and the RHS is her payoff from implementing a_1. The inequality can be written as $\nu_1[X - \frac{1}{2}]sV_2 \geq 0$, which is satisfied applying our argument above.

(iv) Note that (11) is equivalent to

$$\zeta + \lambda(sV_1) + \left[\nu_1 X + (1-\nu_1)\frac{1}{2}\right](sV_2) \geq (1-\lambda)(sV_1) + \left\{\zeta + \frac{1}{2}(sV_2)\right\}, \tag{A.9}$$

where the LHS is her payoff from implementing a_1, and the RHS is her payoff from b_1. **Q.E.D.**

Proof of Proposition 5. **1.** We first write down the candidate equilibrium for this case and then prove that this is indeed an equilibrium:

(i) in period 1, all deputy types send a null report, and the leader adopts the following strategy:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \emptyset\}) \to a_1 \\ (N, \beta) \to b_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to b_1 \end{cases}$$

(ii) in period 1, the principal believes that the deputy must be biased with probability 1 whenever the deputy submits a non-null report, but elsewhere the beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal changes period 1 leader if and only if she implemented b_1, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

That the leader and the deputy's period 1 strategies are optimal follows from Lemma 5, and the fact that (10) holds and (11) is violated. Thus it is sufficient to consider the principal's strategy in period 2. From the leader's strategy note that a_1 is implemented only by an unbiased leader. Thus upon observing a_1, the principal concludes that the leader is honest, and does not replace her in period 2. Next consider the case where b_1 is implemented. From the leader's strategy, if the signal is β, then both a biased and an unbiased leader implements b_1, whereas if the signal is either α, or $\emptyset$, then a biased leader implements b_1, whereas an unbiased leader does not. Thus, on observing b_1, the probability that the leader is unbiased decreases. Given that $\nu_2 < \nu_1$ to begin with, it is optimal to replace her.

2. Consider the following candidate equilibrium:

(i) in period 1, all deputy types send a null report, and the leader adopts the following strategy:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \beta, \emptyset\}) \to a_1 \\ (\neg N, \{\alpha, \beta, \emptyset\}) \to b_1 \end{cases}$$

(ii) in period 1, the principal believes that the deputy must be biased with probability 1 whenever the deputy submits a non-null report, but elsewhere the beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal changes period 1 leader if and only if she implemented b_1, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

That the leader and the deputy's period 1 strategies are optimal follows from Lemma 5, and the fact that (10) and (11) are both violated. Thus it is sufficient to consider the principal's strategy in period 2. From the leader's strategy note that a_1 is implemented only by an unbiased leader, whereas b_1 is implemented only by a biased leader. Thus upon observing a_1, the principal concludes that the leader is honest, and does not replace her in period 2, whereas upon observing b_1 the principal concludes that the leader is biased and replaces her.

3. Consider the following candidate equilibrium:

(i) in period 1, all deputy types send a null report, and the leader adopts the following strategy:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (N, \{\alpha, \beta, \emptyset\}) \rightarrow a_1 \\ (\neg N, \{\beta, \emptyset\}) \rightarrow b_1 \\ (\neg N, \{\alpha\}) \rightarrow a_1 \end{cases}$$

(ii) in period 1, the principal believes that the deputy must be biased with probability 1 whenever the deputy submits a non-null report, but elsewhere the beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal changes period 1 leader if and only if she implemented b_1, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

That the leader and the deputy's period 1 strategies are optimal follows from Lemma 5, and the fact that (10) is violated, and (11) holds. Thus it is sufficient to consider the principal's strategy in period 2. From the leader's strategy note that b_1 is implemented only by a biased leader. Thus upon observing b_1, the principal concludes that the leader is biased, and replaces her in period 2. From the leader's strategy, upon observing a_1 the principal's belief regarding the leader's honesty level increases. However, given that $\frac{2\nu_2}{2\nu_2+(1-\nu_2)(1-\epsilon)} < \nu_1$, she is still replaced.

4. By letting $\epsilon \rightarrow 1$ verify that (10) and (11) are both satisfied: the RHS$\rightarrow 0$ and LHS> 0. So using the fact that the RHS of (10) and (11) is strict decreasing and continuous in ϵ, there exists a unique $\hat{\epsilon} < 1$ such that both (10) and (11) will be satisfied if and only if $\epsilon \geq \hat{\epsilon}$.

At $\epsilon = 0$, RHS of (10)$= (1-\nu_1)\frac{V_2}{2}(2\lambda - 1)$, and RHS of (11)$= \nu_1\frac{V_2}{2}(2\lambda-1)$. So it follows that

$$(2\lambda - 1)V_1 \geq \max\{1-\nu_1, \nu_1\}\frac{V_2}{2}(2\lambda - 1) \Leftrightarrow V_1 \geq \max\{1-\nu_1, \nu_1\}\frac{V_2}{2}$$

will ensure conditions (10) and (11) will be satisfied for all ϵ, given that the RHS of (10) and (11) is strictly decreasing in ϵ.

Now consider the following candidate equilibrium:

(i) in period 1, all deputy types send a null report, and the leader is signal respecting:

$$\mathcal{I}_{2,1}(\lambda) = \begin{cases} (\{N, \neg N\}, \{\alpha\}) \to a_1 \\ (\{N, \neg N\}, \{\beta\}) \to b_1 \\ (\neg N, \{\emptyset\}) \to b_1 \\ (N, \{\emptyset\}) \to a_1 \end{cases}$$

(ii) in period 1, the principal's beliefs are determined by Bayes' rule;

(iii) in period 2,

(a) the principal changes period 1 leader if and only if she implemented b_1, and

(b) period 2 leader and period 2 deputy coordinate on the sincere equilibrium.

That the leader and the deputy's strategies are optimal follows from Lemma 5, and the fact that (10) and (11) both hold. Moreover, note that implementing b_1 (resp. a_1) implies that the leader is biased (resp. honest), so that the principal's strategy is optimal.

5. We prove this result in several steps:

Step 1. From Lemma 4 no other regime-switching rule exists.

Step 2. Next consider a parameter region not covered by Proposition 5.1–5.4: $\frac{2\nu_2}{2\nu_2+(1-\nu_2)(1-\epsilon)} > \nu_1$, (10) is violated and (11) holds. Note that given Lemma 5, the unbiased leader is going to be virtue signalling, and biased leader is going to be signal respecting. In that case, given that $\frac{2\nu_2}{2\nu_2+(1-\nu_2)(1-\epsilon)} > \nu_1$, the leader is not going to be replaced (this mimics the argument in Proposition 5.3). As we have already argued in Lemma 4.1 however, there cannot be an equilibrium where the principal never induces a regime change.

Step 3. Finally, for the parameter zone covered in Proposition 5.1–5.4, Lemma 5 ensures that the leader and the deputy's strategies are unique, so that the only other equilibria that exist must be identical to one of the equilibria in Proposition 5.1–5.4, except for the principal's off-the-equilibrium beliefs. **Q.E.D.**

Proof of Proposition 6. Given that the principal commits to *not* implementing the Yesman-I equilibrium in Proposition 4.3, and that (4) is violated, the Yesman-I equilibrium involves the one in Proposition 4.1(B) if (5) fails and at least one in Proposition 4.1(C) and Proposition 4.2 if (5) holds.

Under Yesman-I, we need to compare the two surpluses from potential multiple equilibria (Proposition 4.1(C) and Proposition 4.2). Comparing these equilibria, the relevant surplus differences are: (12) − (8) and (12) − (9). Below we turn to evaluate these differences.

1. Write (12)–(8)

$$
\begin{aligned}
&= \nu_2\frac{1}{2}\{\epsilon+(1-\epsilon)\lambda\}(s_P V_1)+\nu_2\frac{1}{2}\{\epsilon+(1-\epsilon)\lambda\}X\,(s_P V_2)\\
&\quad+\nu_2\frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\left(\nu_1 X+(1-\nu_1)\frac{1}{2}\right)(s_P V_2)+\nu_2\frac{1}{2}\{\epsilon+(1-\epsilon)(1-\lambda)\}X\,(s_P V_2)\\
&\quad+\nu_2\frac{1}{2}\{(1-\epsilon)\lambda\}(s_p V_1)+\nu_2\frac{1}{2}\{(1-\epsilon)\lambda\}\left(\nu_1 X+(1-\nu_1)\frac{1}{2}\right)(s_P V_2)\\
&+(1-\nu_2)\frac{1}{2}(s_P V_1)+(1-\nu_2)\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)\\
&-\frac{1}{2}\big[\nu_1+(1-\nu_1)\{\epsilon+2(1-\epsilon)\xi\}\big](s_P V_1)-\big[\nu_1 X+(1-\nu_1)\frac{1}{2}\{\frac{1}{2}(1-\epsilon)\}\big](s_P V_2)\\
&\quad-(1-\nu_1)\big[\frac{1}{2}(1+\epsilon)\big(\nu_2 X+(1-\nu_2)\frac{1}{2}\big)\big](s_P V_2)\\
&=\frac{1}{2}\big[\nu_2\{\epsilon+2(1-\epsilon)\lambda\}+(1-\nu_2)-\nu_1-(1-\nu_1)\{\epsilon+2(1-\epsilon)\xi\}\big](s_P V_1)\\
&\quad+\nu_2\frac{1}{2}(1+\epsilon)X\,(s_P V_2)+\nu_2\frac{1}{2}(1-\epsilon)\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)+(1-\nu_2)\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)\\
&\quad-\nu_1 X\,(s_P V_2)-(1-\nu_1)\frac{1}{2}\{\frac{1}{2}(1-\epsilon)\}(s_P V_2)-(1-\nu_1)\frac{1}{2}(1+\epsilon)\big(\nu_2 X+(1-\nu_2)\frac{1}{2}\big)(s_P V_2)\\
&=\frac{1}{2}(1-\epsilon)\big[\nu_2(2\lambda-1)-(1-\nu_1)(2\xi-1)\big](s_P V_1)+[\nu_2\frac{1}{2}(1+\epsilon)-\nu_1]X\,(s_P V_2)\\
&\quad+[\nu_2\frac{1}{2}(1-\epsilon)+(1-\nu_2)]\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)\\
&\quad-(1-\nu_1)\frac{1}{2}\{\frac{1}{2}(1-\epsilon)\}(s_P V_2)-(1-\nu_1)\frac{1}{2}(1+\epsilon)\big(\nu_2 X+(1-\nu_2)\frac{1}{2}\big)(s_P V_2)\\
&=\frac{1}{2}(1-\epsilon)\big[\nu_2(2\lambda-1)-(1-\nu_1)(2\xi-1)\big](s_P V_1)+[\nu_2\frac{1}{2}(1+\epsilon)-\nu_1]\cdot X(s_P V_2)\\
&\quad+[\nu_2\frac{1}{2}(1-\epsilon)+(1-\nu_2)]\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)\\
&\quad-(1-\nu_1)\frac{1}{4}(1-\epsilon)(s_P V_2)-(1-\nu_1)\frac{1}{2}(1+\epsilon)\big(\nu_2 X+(1-\nu_2)\frac{1}{2}\big)(s_P V_2)\\
&=\frac{1}{2}(1-\epsilon)\big[\nu_2(2\lambda-1)-(1-\nu_1)(2\xi-1)\big](s_P V_1)\\
&\quad+\big[\nu_2\frac{1}{2}(1+\epsilon)-\nu_1+\nu_1\nu_2\frac{1}{2}(1-\epsilon)+\nu_1(1-\nu_2)-\nu_2(1-\nu_1)\frac{1}{2}(1+\epsilon)\big]\cdot X\,(s_P V_2)\\
&\quad+(1-\nu_1)\frac{1}{2}\big[\{\nu_2\frac{1}{2}(1-\epsilon)+1-\nu_2\}-\frac{1}{2}(1-\epsilon)-\frac{1}{2}(1+\epsilon)(1-\nu_2)\big](s_P V_2)\\
&=\frac{1}{2}(1-\epsilon)\big[\nu_2(2\lambda-1)-(1-\nu_1)(2\xi-1)\big](s_P V_1)\\
&\quad+\big[-\nu_1+\nu_1\nu_2\frac{1}{2}(1-\epsilon)+\nu_1-\nu_1\nu_2+\nu_1\nu_2\frac{1}{2}(1+\epsilon)\big]\cdot X\,(s_P V_2)\\
&\quad+(1-\nu_1)\frac{1}{2}\big[-\frac{1}{2}(1-\epsilon)(1-\nu_2)+1-\nu_2-\frac{1}{2}(1+\epsilon)(1-\nu_2)\big](s_P V_2)\\
&=\frac{1}{2}(1-\epsilon)\big[\nu_2(2\lambda-1)-(1-\nu_1)(2\xi-1)\big](s_P V_1)\\
&\quad+\big[\nu_1\nu_2-\nu_1\nu_2\big]\cdot X\cdot(s_P V_2)+(1-\nu_1)(1-\nu_2)\frac{1}{4}\big[-(1-\epsilon)+2-(1+\epsilon)\big]\cdot(s_P V_2)\\
&=\frac{1}{2}(1-\epsilon)\big[\nu_2(2\lambda-1)-(1-\nu_1)(2\xi-1)\big](s_P V_1)+0+(1-\nu_1)(1-\nu_2)\frac{1}{4}\cdot 0\\
&=\frac{1}{2}(1-\epsilon)\big[\nu_2(2\lambda-1)-(1-\nu_1)(2\xi-1)\big](s_P V_1),
\end{aligned}
$$

which can be positive or negative depending on the relative worths of the two agents (conditional on equilibrium strategies).

2. Write

$$(12)-(9) = (1-\epsilon)\big[\nu_2(2\lambda-1)-\nu_1(2\xi-1)\big]\frac{1}{2}(s_P V_1)+\nu_2 X\frac{1}{2}(s_P V_2) \\ +\nu_2\epsilon(1-\nu_1)\big[X-\frac{1}{2}\big]\frac{1}{2}(s_P V_2)+\frac{1}{2}\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)\frac{1}{2}(s_P V_2). \tag{A.10}$$

We claim that $(12)-(9)>0$. We will argue by contradiction, so suppose $(12)-(9)\leq 0$.

In what follows below, the first inequality is the negation of (11), whereas the second inequality is $(12)-(9)\leq 0$ (i.e., honesty dominating talent) for which it must be the case that $\big[\nu_2(2\lambda-1)-\nu_1(2\xi-1)\big]<0$ (see (A.10)). Let us now simplify the final right-hand side expression (of the inequalities) in steps:

$$\begin{aligned}
\frac{\nu_1\big[X-\frac{1}{2}\big]}{2\lambda-1} > \frac{V_1}{V_2} \geq \;& \frac{\Big\{\nu_2 X+\nu_2\epsilon(1-\nu_1)\big[X-\frac{1}{2}\big]+\frac{1}{2}\Big(\nu_1 X+(1-\nu_1)\frac{1}{2}\Big)\Big\}}{(1-\epsilon)\underbrace{\Big[\nu_1(2\xi-1)-\nu_2(2\lambda-1)\Big]}_{>0 \text{ as implied by } (12)-(9)\leq 0}} \\
=\;& \frac{\nu_2 X+[\nu_2\epsilon(1-\nu_1)+\frac{1}{2}\nu_1]\big[X-\frac{1}{2}\big]+\frac{1}{4}}{(1-\epsilon)\big[\nu_1(2\xi-1)-\nu_2(2\lambda-1)\big]} \\
=\;& \frac{\nu_2\big[X-\frac{1}{2}\big]+[\nu_2\epsilon(1-\nu_1)+\frac{1}{2}\nu_1]\big[X-\frac{1}{2}\big]+\frac{1}{2}(\frac{1}{2}+\nu_2)}{(1-\epsilon)\big[\nu_1(2\xi-1)-\nu_2(2\lambda-1)\big]} \\
=\;& \frac{[\nu_2\epsilon(1-\nu_1)+\frac{1}{2}\nu_1+\nu_2]\big[X-\frac{1}{2}\big]+\frac{1}{2}(\frac{1}{2}+\nu_2)}{(1-\epsilon)\big[\nu_1(2\xi-1)-\nu_2(2\lambda-1)\big]} \\
>\;& \frac{[\nu_2\epsilon(1-\nu_1)+\frac{1}{2}\nu_1+\nu_2]\big[X-\frac{1}{2}\big]+\big[X-\frac{1}{2}\big](\frac{1}{2}+\nu_2)}{(1-\epsilon)\big[\nu_1(2\xi-1)-\nu_2(2\lambda-1)\big]} \\
=\;& \frac{[\nu_2\epsilon(1-\nu_1)+\frac{1}{2}\nu_1+\frac{1}{2}+2\nu_2]\big[X-\frac{1}{2}\big]}{(1-\epsilon)\big[\nu_1(2\xi-1)-\nu_2(2\lambda-1)\big]}.
\end{aligned}$$

This leads to a contradiction because (i) the numerator of the most right-hand side expression exceeds the numerator of the most left-hand side expression ($\frac{1}{2}\nu_1+\frac{1}{2}\cdot 1\geq \nu_1$), and (ii) the denominator of the most right-hand side is strictly less than the denominator of the left-hand side. The latter can be seen by combining

$$2\lambda-1>\nu_1(2\xi-1)-\nu_2(2\lambda-1)\Leftrightarrow(1+\nu_2)(2\lambda-1)>\nu_1(2\xi-1)$$

(which is true because $1+\nu_2>\nu_1$ and $\lambda>\xi$), and the fact that $1-\epsilon\leq 1$.

3. Write $(12)-(7)$

$$\begin{aligned}
&=\nu_2\frac{1}{2}\{\epsilon+(1-\epsilon)\lambda\}(s_P V_1)+\nu_2\frac{1}{2}\{\epsilon+(1-\epsilon)\lambda\}X(s_P V_2) \\
&\quad+\nu_2\frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)+\nu_2\frac{1}{2}\{\epsilon+(1-\epsilon)(1-\lambda)\}X(s_P V_2) \\
&\quad+\nu_2\frac{1}{2}\{(1-\epsilon)\lambda\}(s_p V_1)+\nu_2\frac{1}{2}\{(1-\epsilon)\lambda\}\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2) \\
&+(1-\nu_2)\frac{1}{2}(s_P V_1)+(1-\nu_2)\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{2}(s_P V_1) - \nu_1 X\,(s_P V_2) - (1-\nu_1)\big(\nu_2 X + (1-\nu_2)\frac{1}{2}\big)(s_P V_2)\\
&= \frac{1}{2}\Big\{\nu_2(1-\epsilon)(2\lambda-1)\Big\}(s_P V_1) + \nu_2\frac{1}{2}\Big[(1+\epsilon)X + (1-\epsilon)\big(\nu_1 X + (1-\nu_1)\frac{1}{2}\big)\Big](s_P V_2)\\
&+ (1-\nu_2)\big(\nu_1 X + \cancel{(1-\nu_1)\frac{1}{2}}\big)(s_P V_2) - \nu_1 X(s_P V_2) - (1-\nu_1)\big[\nu_2 X + \cancel{(1-\nu_2)\frac{1}{2}}\big](s_P V_2)\\
&= \frac{1}{2}\{\nu_2(1-\epsilon)(2\lambda-1)\}(s_P V_1) + [\nu_2\frac{1}{2}(1+\epsilon) - \nu_1]X\,(s_P V_2)\\
&\qquad + \nu_2\frac{1}{2}(1-\epsilon)\big(\nu_1 X + (1-\nu_1)\frac{1}{2}\big)(s_P V_2) + (1-\nu_2)\nu_1 X\,(s_P V_2) - (1-\nu_1)\nu_2\, X(s_P V_2)\\
&= \frac{1}{2}\{\nu_2(1-\epsilon)(2\lambda-1)\}(s_P V_1) + [\nu_2\frac{1}{2}(1+\epsilon) - \cancel{\nu_1}]X\,(s_P V_2)\\
&\qquad + \nu_2\frac{1}{2}(1-\epsilon)\big(\nu_1 X + (1-\nu_1)\frac{1}{2}\big)(s_P V_2) + (\cancel{\nu_1} - \nu_2)X\,(s_P V_2)\\
&= \frac{1}{2}\{\nu_2(1-\epsilon)(2\lambda-1)\}(s_P V_1) - \nu_2\frac{1}{2}(1-\epsilon)X\,(s_P V_2) + \nu_2\frac{1}{2}(1-\epsilon)\big(\nu_1 X + (1-\nu_1)\frac{1}{2}\big)(s_P V_2)\\
&= \frac{1}{2}\{\nu_2(1-\epsilon)(2\lambda-1)\}(s_P V_1) - (1-\nu_1)\nu_2\frac{1}{2}(1-\epsilon)X\,(s_P V_2) + \big(\nu_2\frac{1}{2}(1-\epsilon)(1-\nu_1)\frac{1}{2}\big)(s_P V_2)\\
&= \frac{1}{2}\{\nu_2(1-\epsilon)(2\lambda-1)\}(s_P V_1) + (1-\nu_1)\nu_2\frac{1}{2}(1-\epsilon)\big[1-X\big](s_P V_2) > 0,
\end{aligned}
$$

since $\big[1-X\big] > 0$. **Q.E.D.**

Proof of Proposition 7. **1.** Write (8) − (13)

$$
\begin{aligned}
&= \frac{1}{2}\Big[\nu_1 + (1-\nu_1)\big\{\epsilon + 2(1-\epsilon)\xi\big\}\Big](s_P V_1)\\
&+ \Big[\nu_1\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\big\} + (1-\nu_1)\frac{1}{2}\big\{\frac{1}{2}(1-\epsilon)\big\}\Big](s_P V_2)\\
&+ (1-\nu_1)\Big[\frac{1}{2}(1+\epsilon)\Big(\nu_2\big[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\big] + (1-\nu_2)\frac{1}{2}\Big)\Big](s_P V_2)\\
&- \frac{1}{2}(s_P V_1) - \nu_2\Big[\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\big\}(s_P V_2)\Big]\\
&- (1-\nu_2)\Big[\Big(\nu_1\big[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\big] + (1-\nu_1)\frac{1}{2}\Big)(s_P V_2)\Big]\\
&= (1-\nu_1)(1-\epsilon)\Big[\xi - \frac{1}{2}\Big](s_P V_1) + \Big[\nu_1 X + (1-\nu_1)\frac{1}{2}\big\{\frac{1}{2}(1-\epsilon)\big\}\Big](s_P V_2)\\
&+ (1-\nu_1)\Big[\frac{1}{2}(1+\epsilon)\big\{\nu_2 X + (1-\nu_2)\frac{1}{2}\big\}\Big](s_P V_2)\\
&- \nu_2 X\,(s_P V_2) - (1-\nu_2)\big[\nu_1 X + (1-\nu_1)\frac{1}{2}\big](s_P V_2)\\
&= (1-\nu_1)(1-\epsilon)\Big[\xi - \frac{1}{2}\Big](s_P V_1) + (\nu_1 - \nu_2)X\,(s_P V_2)\\
&+ (1-\nu_1)\Big[\frac{1}{4}(1-\epsilon) + \frac{1}{2}(1+\epsilon)\big\{\nu_2 X + (1-\nu_2)\frac{1}{2}\big\}\Big](s_P V_2) - (1-\nu_2)\big[\nu_1 X + (1-\nu_1)\frac{1}{2}\big](s_P V_2)\\
&= (1-\nu_1)(1-\epsilon)\Big[\xi - \frac{1}{2}\Big](s_P V_1) + \Big[(\nu_1 - \nu_2) + \frac{1}{2}(1+\epsilon)(1-\nu_1)\nu_2 - (1-\nu_2)\nu_1\Big]X\cdot(s_P V_2)\\
&+ (1-\nu_1)\Big[\frac{1}{4}(1-\epsilon) + \frac{1}{4}(1+\epsilon)(1-\nu_2) - \frac{1}{2}(1-\nu_2)\Big](s_P V_2)\\
&= (1-\nu_1)(1-\epsilon)\Big[\xi - \frac{1}{2}\Big](s_P V_1) - \frac{1}{2}(1-\epsilon)(1-\nu_1)\nu_2 X\,(s_P V_2) + (1-\nu_1)\nu_2\frac{1}{4}(1-\epsilon)(s_P V_2)
\end{aligned}
$$

$$= (1-\nu_1)(1-\epsilon)\left[\xi - \frac{1}{2}\right](s_P V_1) - \frac{1}{2}(1-\epsilon)(1-\nu_1)\nu_2 \underbrace{[X - \frac{1}{2}]}_{>0}(s_P V_2) > 0,$$

for $\epsilon < 1, \nu_1 < 1$ and $\nu_2 > 0$, because (5) holds.

2. Write (15) − (8)

$$\begin{aligned}
&= \Big[\frac{1}{2}\{(1-\epsilon)\lambda\}\Big\{s_P V_1 + \Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)s_P V_2\Big\} + \frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\Big\{\Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\} \\
&\quad + \frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\Big\{\Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)s_P V_2\Big\} + \frac{1}{2}\{(1-\epsilon)\lambda\}\Big\{s_p V_1 + \Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\}\Big] \\
&+ (1-\nu_2)\Big[\epsilon\Big\{\frac{1}{2}s_P V_1 + \Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)s_P V_2\Big\}\Big] + \nu_2\Big[\epsilon\Big\{\frac{1}{2}s_P V_1 + \Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)s_P V_2\Big\}\Big] \\
&- \frac{1}{2}\Big[\nu_1 + (1-\nu_1)\{\epsilon + 2(1-\epsilon)\xi\}\Big](s_P V_1) - \Big[\nu_1 X + (1-\nu_1)\frac{1}{2}\{\frac{1}{2}(1-\epsilon)\}\Big](s_P V_2) \\
&\quad - (1-\nu_1)\Big[\frac{1}{2}(1+\epsilon)\Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)\Big](s_P V_2) \\
&= (1-\epsilon)\Big[(\lambda - \frac{1}{2}) - (1-\nu_1)(\xi - \frac{1}{2})\Big](s_P V_1) + \frac{1}{2}(1-\epsilon)\Big[(\nu_1+\nu_2)X + (1-\nu_1+1-\nu_2)\frac{1}{2}\Big](s_P V_2) \\
&\quad + (1-\nu_2)\epsilon\Big(\nu_1 X + (1-\nu_1)\frac{1}{2}\Big)(s_P V_2) + \nu_2\epsilon\Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)(s_P V_2) \\
&\quad - \Big[\nu_1 X + (1-\nu_1)\frac{1}{2}\{\frac{1}{2}(1-\epsilon)\}\Big](s_P V_2) - (1-\nu_1)\Big[\frac{1}{2}(1+\epsilon)\Big(\nu_2 X + (1-\nu_2)\frac{1}{2}\Big)\Big](s_P V_2),
\end{aligned}$$

which can be positive or negative; see the second panel of Fig. 6. **Q.E.D.**

Proof of Lemma 6. Write (12) − (6)

$$\begin{aligned}
&= \nu_2\Big[\frac{1}{2}\{\epsilon + (1-\epsilon)\lambda\}\{s_P V_1 + X(s_P V_2)\} + \frac{1}{2}\{(1-\epsilon)(1-\lambda)\}\{(\nu_1 X + (1-\nu_1)\frac{1}{2})s_P V_2\} \\
&\qquad + \frac{1}{2}\{\epsilon + (1-\epsilon)(1-\lambda)\}\{X(s_P V_2)\} + \frac{1}{2}\{(1-\epsilon)\lambda\}\{s_p V_1 + (\nu_1 X + (1-\nu_1)\frac{1}{2})s_P V_2\}\Big] \\
&\quad + (1-\nu_2)\Big[\frac{1}{2}(s_P V_1) + (\nu_1 X + (1-\nu_1)\frac{1}{2})s_P V_2\Big] - \frac{1}{2}[\nu_1\{\epsilon + 2(1-\epsilon)\xi\} + (1-\nu_1)](s_P V_1) \\
&\quad - [\frac{1}{2}\nu_1(1-\epsilon) + (1-\nu_1)](\nu_2 X + (1-\nu_2)\frac{1}{2})(s_P V_2) - \frac{1}{2}\nu_1(1+\epsilon)X(s_P V_2) \\
&= \Big([\nu_2\frac{1}{2}\{\epsilon + 2(1-\epsilon)\lambda\} + (1-\nu_2)\frac{1}{2}] - [\nu_1\frac{1}{2}\{\epsilon + 2(1-\epsilon)\xi\} + (1-\nu_1)\frac{1}{2}]\Big)(s_P V_1) \\
&\quad + \nu_2\Big[\frac{1}{2}\{\epsilon + (1-\epsilon)\lambda\}X(s_P V_2) + \frac{1}{2}\{(1-\epsilon)(1-\lambda)\}(\nu_1 X + (1-\nu_1)\frac{1}{2})(s_P V_2) \\
&\qquad + \frac{1}{2}\{\epsilon + (1-\epsilon)(1-\lambda)\}X(s_P V_2) + \frac{1}{2}\{(1-\epsilon)\lambda\}(\nu_1 X + (1-\nu_1)\frac{1}{2})(s_P V_2)\Big] \\
&\quad + (1-\nu_2)(\nu_1 X + (1-\nu_1)\frac{1}{2})(s_P V_2) - [\frac{1}{2}\nu_1(1-\epsilon) + (1-\nu_1)](\nu_2 X + (1-\nu_2)\frac{1}{2})(s_P V_2) \\
&\qquad - \frac{1}{2}\nu_1(1+\epsilon)X(s_P V_2)
\end{aligned}$$

$$
\begin{aligned}
&= (1-\epsilon)\frac{1}{2}\Big[\nu_2(2\lambda-1)-\nu_1(2\xi-1)\Big](s_P V_1)\\
&\quad + \nu_2\Big[\frac{1}{2}(1+\epsilon)X(s_P V_2)+\frac{1}{2}(1-\epsilon)\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)\Big] + (1-\nu_2)\big(\nu_1 X+(1-\nu_1)\frac{1}{2}\big)(s_P V_2)\\
&\quad - \big[\frac{1}{2}\nu_1(1-\epsilon)+(1-\nu_1)\big]\big(\nu_2 X+(1-\nu_2)\frac{1}{2}\big)(s_P V_2) - \frac{1}{2}\nu_1(1+\epsilon)X\,(s_P V_2)\\
&= (1-\epsilon)\frac{1}{2}\Big[\nu_2(2\lambda-1)-\nu_1(2\xi-1)\Big](s_P V_1) - (\nu_1-\nu_2)\big[\frac{1}{2}(1+\epsilon)\big]X\,(s_P V_2)\\
&\quad + \frac{1}{2}(1-\epsilon)\Big(\cancel{\nu_2\nu_1 X}+\nu_2(1-\nu_1)\frac{1}{2}\Big)(s_P V_2) + (1-\nu_2)\Big(\nu_1 X+\cancel{(1-\nu_1)\frac{1}{2}}\Big)(s_P V_2)\\
&\quad - \frac{1}{2}(1-\epsilon)\Big(\cancel{\nu_1\nu_2 X}+\nu_1(1-\nu_2)\frac{1}{2}\Big)(s_P V_2) - (1-\nu_1)\Big(\nu_2 X+\cancel{(1-\nu_2)\frac{1}{2}}\Big)(s_P V_2)\\
&= (1-\epsilon)\frac{1}{2}\Big[\nu_2(2\lambda-1)-\nu_1(2\xi-1)\Big](s_P V_1) - (\nu_1-\nu_2)\big[\frac{1}{2}(1+\epsilon)\big]X\,(s_P V_2)\\
&\quad - \frac{1}{2}(1-\epsilon)(\nu_1-\nu_2)\frac{1}{2}(s_P V_2) + (\nu_1-\nu_2)X(s_P V_2)\\
&= (1-\epsilon)\frac{1}{2}\Big[\nu_2(2\lambda-1)-\nu_1(2\xi-1)\Big](s_P V_1) + \frac{1}{2}(1-\epsilon)(\nu_1-\nu_2)\big[X-\frac{1}{2}\big](s_P V_2),
\end{aligned}
$$

which can be positive or negative. See Fig. 7. **Q.E.D.**

Proof of Proposition 8. Given that $\frac{\nu_1(1-\epsilon)}{2-\nu_1(1+\epsilon)} < \nu_2$, (4) holds, and (5) is violated, Yesman-I equilibrium in Proposition 4.1(A) exists. Moreover, from Proposition 4.2 and 4.3, no other Yesman-I equilibrium exists. Thus, the principal compares (6), which is his payoff under Yesman-I equilibrium in Proposition 4.1(A), with (12), his payoff under Yesman-II equilibrium in Proposition 5.1. From Lemma 6 we have:

$$
(12)-(6) = (1-\epsilon)\frac{1}{2}\Big[\big\{\nu_2(2\lambda-1)-\nu_1(2\xi-1)\big\}(s_P V_1)+(\nu_1-\nu_2)\big(X-\frac{1}{2}\big)(s_P V_2)\Big]. \qquad \text{(A.11)}
$$

1. Note that $X-\frac{1}{2}>0$. Thus agent 2 having a higher net worth, i.e. $\nu_2(2\lambda-1)-\nu_1(2\xi-1)>0$, is a sufficient condition for (12) − (6) to be positive.

2. The result follows since in this case (i) $\nu_2(2\lambda-1)-\nu_1(2\xi-1)<0$, and (ii) $[(12)-(6)]|_{\epsilon=1}<0$. **Q.E.D.**

References

Agastya, M., Bag, P.K. and Chakraborty, I. (2014). Communication and Authority with a Partially Informed Expert. *Rand Journal of Economics*, 45(1), 176–197.

Aghion, P. and Jackson, M.O. (2016). Inducing Leaders to Take Risky Decisions: Dismissal, Tenure, and Term Limits. *American Economic Journal: Microeconomics*, 8(3), 1–38.

Aghion, P. and Tirole, J. (1997). Formal and Real Authority in Organizations. *Journal of Political Economy*, 105, 1–29.

Almas, I., Cappelen, A.W. and Tungodden, B. (2020). Cutthroat Capitalism versus Cuddly Socialism: Are Americans More Meritocratic and Efficiency-Seeking than Scandinavians? *Journal of Political Economy*, 128(5), 1753–1788.

Alonso, R. and Matouschek, N. (2007). Relational Delegation. *Rand Journal of Economics*, 38, 1070–1089.

Alonso, R. and Matouschek, N. (2008). Optimal Delegation. *Review of Economic Studies*, 75, 259–293.

Armstrong, M. and Vickers, J. (2010). A Model of Delegated Project Choice. *Econometrica*, 78, 213–244.

Arrow, K., Bowles, S. and Durlauf, S.N. (Editors). (2000). *Meritocracy and Economic Inequality*, Princeton University Press, Princeton, New Jersey.

Bag, P. and Roy Chowdhury, P. (2021). Honesty or Talent in Project Implementation? *Working paper* (retired), available on request.

Bag, P. and Roy Chowdhury, P. (2022). Dynamic Delegation. *Working paper.*

Bag, P.K. and Sharma, T. (2019). Sequential Expert Advice: Superiority of Closed-Door Meetings. *International Economic Review*, 60, 1877–1910.

Baker, G., Gibbons, R. and Murphy, K.J. (1999). Informal Authority in Organizations. *Journal of Law, Economics, & Organization*, 15(1), 56–73.

Benabou, R. and Laroque, G. (1992). Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility. *Quarterly Journal of Economics*, 107, 921–958.

Bester H. and Krahmer, D. (2008). Delegation and Incentives. *Rand Journal of Economics*, 39(3), 664-682.

Birch, S. and Allen, N. (2010). How Honest Do Politicians Need to Be? *The Political Quarterly*, 81(1), 49–56.

Blanes i Vidal, J. and Moller, M. (2016). Project Selection and Execution in Teams. *Rand Journal of Economics*, 47, 166–185.

Burgess, R. (2020). A Political History of Backstabbing. https://www.americanpurpose.com/articles/a-political-history-of-backstabbing/.

Burguet, R. and Che, Y-K. (2004). Competitive Procurement with Corruption. *Rand Journal of Economics*, 35(1), 50–68.

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

Compte, O., Lambert-Mogiliansky, A. and Verdier, T. (2005). Corruption and Competition in Procurement Auctions. *Rand Journal of Economics*, 36(1), 1–15.

Crawford, V.P. and Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, 50(6), 1431–1450.

Dessein, W. (2002). Authority and Communication in Organizations. *Review of Economic Studies*, 69(4), 811–838.

Dewatripont, M. and Tirole, J. (1999). Advocates. *Journal of Political Economy*, 107, 1, 1–39.

Detert, J.R. and Edmondson, A.C. (2007). Why Employees Are Afraid to Speak. *Harvard Business Review*, 85(5), 23–25.

Diermeier, D., Keane, M. and Merlo, A. (2005). A Political Economy Model of Congressional Careers. *American Economic Review*, 95, 347–373.

The Economist (2009). Pants on Fire. July 9 report (Bagehot). https://www.economist.com/britain/2009/07/09/pants-on-fire

Ely, J. and Valimaki, J. (2003). Bad Reputation. *Quarterly Journal of Economics*, 118, 785–814.

Fehr, E., Herz, H. and Wilkening, T. (2013). The Lure of Authority: Motivation and Incentive Effects of Power. *American Economic Review*, 103(4), 1325–1359.

Fisman, R. (2001). Estimating the Value of Political Connections. *American Economic Review*, 91(4), 1095–1102.

Golosov, M., Skreta, V., Tsyvinski, A. and Wilson, A. (2014). Dynamic Strategic Information Transmission. *Journal of Economic Theory*, 151, 304-–341.

Holmstrom, B. (1984). On the Theory of Delegation. in M. Boyer and R. Kihlstrom (eds.) *Bayesian Models in Economic Theory*, New York: North-Holland.

Issacharoff, S. (2010). On Political Corruption. *Harvard Law Review*, 124(1), 118–142.

Kartik, N. (2009). Strategic Communication with Lying Costs. *Review of Economic Studies*, 76, 1359–1395.

Koessler, F. and Martimort, D. (2012). Optimal Delegation with Multi-Dimensional Decisions. *Journal of Economic Theory*, 147, 1850–1881.

Konrad, K.A. (2000). Sabotage in Rent-seeking Contests. *Journal of Law, Economics, & Organization*, 16(1), 155–165.

Krahmer, D. (2008). Message-Contingent Delegation. *Journal of Economic Behavior and Organization*, 60, 490-506.

Krishna, V. and Morgan, J. (2001). A Model of Expertise. *Quarterly Journal of Economics*, 116(2), 747–775.

Levy, G. (2007). Decision Making in Committees: Transparency, Reputation and Voting Rules. *American Economic Review*, 97, 150–168.

Li, J., Matouschek, N. and Powell, M. (2017). Power Dynamics in Organizations. *American Economic Journal: Microeconomics*, 9(1), 217–241.

Malone, P. and Hayes, J. (2012). Backstabbing in Organizations: Employees' Perceptions of Incidents, Motives, and Communicative Responses. *Communication Studies*, 63(2), 194-219, DOI: 10.1080/10510974.2011.635552.

Mironov, M. (2013). Taxes, Theft, and Firm Performance. *Journal of Finance*, vol. LXVIII, No. 4, 1441–1472.

Moisson, P-H and Tirole, J. (2020). Cooptation: Meritocracy vs. Homophily in Organizations. Available at https://www.tse-fr.eu/sites/default/files/TSE/documents/doc/by/tirole/cooptation_2020_05_15.pdf

Morris, S. (2001). Political Correctness. *Journal of Political Economy*, 109(2), 231–265.

O'Neill, O. (2002). *A Question of Trust: The BBC Reith Lectures*, Cambridge, Cambridge University Press.

Ottaviani, M. and Sorensen, P.N. (2006). Professional Advice. *Journal of Economic Theory*, 126, 120–142.

Prendergast, C. (1993). A Theory of "Yes Men". *American Economic Review*, 83(4), 757–770.

Shin, H.S. (1998). Adversarial and Inquisitorial Procedures in Arbitration. *Rand Journal of Economics*, 29, 2, 378–405.

Smith, J.D. (2016). US Political Corruption and Firm Financial Policies. *Journal of Financial Economics*, 121(2), 350–367.

Sobel, J. (1985). A Theory of Credibility. *Review of Economic Studies*, 52, 557–573.

Tirole, J. (2006). *The Theory of Corporate Finance.* Publisher: Princeton University Press, ISBN 9780691125565.

Troy, T. (2020). *Fight House: Rivalries in the White House from Truman to Trump.* Publisher: Regnery History, ISBN 1621578364.

B Supplementary material: Informative Advice

Fixing sincere equilibrium in period 2, we step back to analyze advice and project implementation decisions in period 1. Initially we study a truth-telling equilibrium involving deputy's advice. Project implementation by the leader will not distort from ex-ante efficiency despite the lure of corruption or the pressure to appear as politically correct by straying from choices that might be deemed as ill motivated.

■ **Continuation game starting at T_{12}.** Letting $\ell = 1$ in period 1 who is known to be of talent ξ and $d = 2$ known to be of talent λ, we shall focus on what we call an *Informative advice equilibrium* ($\mathcal{E}_1$) in which a talented deputy reveals her signal correctly. This equilibrium is 'simple' in that the posterior beliefs (about the state or the deputy's bias), and the project implementation or regime-switching decisions do not depend on the strategic manipulations by the deputy. We begin by delineating the strategies in this candidate equilibrium.

Period 1. *Stage* T_{12}. The λ-talent *deputy's* strategies are sincere and thus pooling with respect to her bias type:

$$R_1(\lambda) = \begin{cases} (\{N, \neg N\}, \alpha) \to \hat{\alpha} \\ (\{N, \neg N\}, \beta) \to \hat{\beta} \\ (\{N, \neg N\}, \emptyset) \to \hat{\emptyset} \end{cases} \tag{B.1}$$

Stage T_{13}. The ξ-talent *leader* implements decisions according to what the deputy recommends if the report is non-null, and otherwise follows her own signal except when she also fails to observe a signal. In this last scenario she implements a_1 when biased and b_1 when unbiased. That is,

$$\mathcal{I}_{1,1}(\xi) = \begin{cases} (\{N, \neg N\}, \{\alpha, \beta, \emptyset\}) \times (\hat{\alpha}) \to a_1 \\ (\{N, \neg N\}, \alpha) \times (\hat{\emptyset}) \to a_1 \\ (\neg N, \emptyset) \times (\hat{\emptyset}) \to a_1 \\ (\{N, \neg N\}, \{\alpha, \beta, \emptyset\}) \times (\hat{\beta}) \to b_1 \\ (\{N, \neg N\}, \beta) \times (\hat{\emptyset}) \to b_1 \\ (N, \emptyset) \times (\hat{\emptyset}) \to b_1 \end{cases} \tag{B.2}$$

Period 2. *Stage* T_{21}. The *principal's* strategy depends on whether the deputy makes a null recommendation or not.

If the deputy makes a null recommendation, then the principal's strategy is to replace period 1 leader if she implements the project favored by her corrupt self, i.e., a_1, and otherwise retain the leader for period 2.

If the deputy makes a non-null recommendation, then there is regime change if and only if the leader implements a project that goes against the deputy's recommendation.

Stage T_{22}. The *deputy's* recommendation strategies are same as the one by period 1 deputy.

Stage T_{23}. The *leader's* project choice follows the sincere equilibrium strategies as in Proposition 1.

Beliefs: The posterior beliefs about the deputy's types will be determined using Bayes' rule. There is no out-of-equilibrium report by the deputy in period 1 or period 2. ||

Note that while communication by the deputy is cheap talk, the leader's action can be costly as it may involve sacrifice of current bias benefits.

Proposition 11 ***(*Corruption vs. reputation*)*** *Let agent 1, who is of lesser talent ξ, be the period 1 leader, and fix the principal's and the talented (λ) deputy's strategies as in $\mathcal{E}_1$. Then the informative advice equilibrium $\mathcal{E}_1$ of the Stage T_{12} game exhibits the following properties:*

(i) If the deputy makes a recommendation a_1 or b_1 in period 1 by reporting $\hat{\alpha}$ or $\hat{\beta}$, then the leader implements the recommended project under condition (B.7) derived in the proof.

(ii) *Suppose the deputy submits a null recommendation ($\hat{\emptyset}$). Then under conditions* (B.8) *and* (B.10) *derived in the proof, the following hold:*

(a) *the leader implements the project suggested by her own non-null signal even if it means being ousted, whereas*

(b) *if she observes a null ($\emptyset$) signal then she implements her pet project a_1 and lose leadership if* biased *and implements b_1 to avoid being ousted if* unbiased.

Proof of Proposition 11. (i) Consider the leader's decisions facing a clear recommendation, a_1 or b_1.

Clearly, any deviation by an unbiased leader either lowers her period 1 payoff (as she goes against deputy's strong signal when her signal is weak and opposite to that of the deputy or has null signal), or lowers her second period payoff due to loss of leadership, or both.

So suppose the leader is biased. First consider the case where the deputy reports $\hat{\beta}$.

Case 1: leader has α signal. The leader's payoff from implementing project b_1 equals

$$\Pr(b_1|(\alpha,\beta),(\xi,\lambda))\cdot(sV_1)+[\zeta+\frac{1}{2}(sV_2)]=\frac{(1-\xi)\lambda}{(1-\xi)\lambda+\xi(1-\lambda)}(sV_1)+[\zeta+\frac{1}{2}(sV_2)],$$

whereas her payoff from deviation to a_1 is

$$\zeta+\frac{\xi(1-\lambda)}{(1-\xi)\lambda+\xi(1-\lambda)}(sV_1)+\Big[\nu_2\big\{(1-\epsilon)\lambda+\epsilon(1-\epsilon)\xi+\epsilon^2(\frac{1}{2})\big\}(sV_2)+(1-\nu_2)\frac{1}{2}(sV_2)\Big].$$

So the leader would prefer implementing b_1 if

$$\frac{\lambda-\xi}{(1-\xi)\lambda+\xi(1-\lambda)}(sV_1)\geq\nu_2\Big[\{(1-\epsilon)\lambda+\epsilon(1-\epsilon)\xi+\epsilon^2(\frac{1}{2})\}-\frac{1}{2}\Big]sV_2. \tag{B.3}$$

Case 2: leader has β signal. The leader's payoff from implementing b_1 equals

$$\Pr(b_1|(\beta,\beta),(\xi,\lambda))\cdot(sV_1)+[\zeta+\frac{1}{2}(sV_2)]=\frac{\xi\lambda}{\xi\lambda+(1-\xi)(1-\lambda)}(sV_1)+\big[\zeta+\frac{1}{2}(sV_2)\big],$$

whereas her payoff from deviation to a_1 is

$$\zeta+[1-\frac{\xi\lambda}{\xi\lambda+(1-\xi)(1-\lambda)}](sV_1)+\Big[\nu_2\big\{(1-\epsilon)\lambda+\epsilon(1-\epsilon)\xi+\epsilon^2(\frac{1}{2})\big\}sV_2+(1-\nu_2)\frac{1}{2}(sV_2)\Big].$$

So the leader would prefer implementing b_1 if

$$\frac{\lambda+\xi-1}{\xi\lambda+(1-\xi)(1-\lambda)}(sV_1)\geq\nu_2\Big[\{(1-\epsilon)\lambda+\epsilon(1-\epsilon)\xi+\epsilon^2(\frac{1}{2})\}-\frac{1}{2}\Big]sV_2. \tag{B.4}$$

Case 3: leader has $\emptyset$ signal. The leader's payoff from implementing b_1 equals

$$\begin{aligned}&\Pr(b_1|(\emptyset,\beta),(\xi,\lambda))\cdot(sV_1)+\big[\zeta+\frac{1}{2}(sV_2)\big]\\&=\frac{\Pr(\emptyset,\beta|b_1,\xi,\lambda)\Pr(b_1)}{\Pr(\emptyset,\beta|b_1,\xi,\lambda)\Pr(b_1)+\Pr(\emptyset,\beta|a_1,\xi,\lambda)\Pr(a_1)}\cdot(sV_1)+\big[\zeta+\frac{1}{2}(sV_2)\big]\\&=\frac{\epsilon(1-\epsilon)\lambda(1/2)}{\epsilon(1-\epsilon)\lambda(1/2)+\epsilon(1-\epsilon)(1-\lambda)(1/2)}(sV_1)+\big[\zeta+\frac{1}{2}(sV_2)\big]\\&=\lambda(sV_1)+\big[\zeta+\frac{1}{2}(sV_2)\big],\end{aligned}$$

whereas her payoff from deviation to a_1 is

$$\zeta + (1-\lambda)(sV_1) + \big[\nu_2\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\}sV_2 + (1-\nu_2)\frac{1}{2}(sV_2)\big].$$

So the leader would prefer implementing b_1 if

$$(2\lambda-1)(sV_1) \geq \nu_2\left[\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\} - \frac{1}{2}\right]sV_2. \tag{B.5}$$

Summary: Combining (B.3)–(B.5), low-talent, biased leader facing $\hat{\beta}$ report from a talented deputy will not deviate from her equilibrium project implementation decision if

$$\begin{aligned}\min\{2\lambda-1, \frac{\lambda+\xi-1}{\xi\lambda+(1-\xi)(1-\lambda)}, \frac{\lambda-\xi}{(1-\xi)\lambda+\xi(1-\lambda)}\}(sV_1)\\ \geq \nu_2\left[\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\} - \frac{1}{2}\right]sV_2.\end{aligned} \tag{B.6}$$

In a bid to simplify (B.6), we first show that $2\lambda - 1 \leq \frac{\lambda+\xi-1}{\xi\lambda+(1-\lambda)(1-\xi)}$. Note that

$$\begin{aligned}2\lambda-1 \leq \frac{\lambda+\xi-1}{\xi\lambda+(1-\lambda)(1-\xi)} &\Leftrightarrow 2\lambda[\xi\lambda+(1-\lambda)(1-\xi)] - \xi\lambda - 1 + \xi + \lambda - \xi\lambda \leq \lambda+\xi-1\\ &\Leftrightarrow 2\lambda[\xi\lambda+(1-\lambda)(1-\xi)] \leq 2\lambda\xi \leftrightarrow \xi\lambda+(1-\lambda)(1-\xi) \leq \xi\\ &\Leftrightarrow (1-\xi)(1-\lambda) \leq \xi(1-\lambda) \leftrightarrow 1 \leq 2\xi,\end{aligned}$$

which holds as $\xi \geq \frac{1}{2}$.

We then show that $\frac{\lambda-\xi}{(1-\xi)\lambda+\xi(1-\lambda)} \leq 2\lambda-1$. Note that

$$\begin{aligned}\frac{\lambda-\xi}{(1-\xi)\lambda+\xi(1-\lambda)} \leq 2\lambda-1 &\Leftrightarrow \lambda-\xi \leq -\lambda-\xi+2\lambda\xi+2\lambda[(1-\xi)\lambda+\xi(1-\lambda)]\\ &\Leftrightarrow 2\lambda(1-\xi) \leq 2\lambda[(1-\xi)\lambda+\xi(1-\lambda)] \leftrightarrow 1-\xi \leq (1-\xi)\lambda+\xi(1-\lambda)\\ &\Leftrightarrow (1-\xi)(1-\lambda) \leq \xi(1-\lambda) \leftrightarrow 1 \leq 2\xi,\end{aligned}$$

which holds since $\xi \geq \frac{1}{2}$.

Consequently the LHS of (B.6) simplifies to $\frac{\lambda-\xi}{(1-\xi)\lambda+\xi(1-\lambda)}(sV_1)$. Hence (B.6) simplifies to a single inequality

$$\frac{\lambda-\xi}{(1-\xi)\lambda+\xi(1-\lambda)}(sV_1) \geq \nu_2\left[\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\} - \frac{1}{2}\right]sV_2. \tag{B.7}$$

Finally, let a biased leader face a report $\hat{\alpha}$. By playing her equilibrium strategy a_1, the leader implements ex-ante efficient decision in period 1 and obtains her bias payoff, thereby maximizing period 1 payoff, and retains leadership. Thus there is no incentive to deviate. This completes the proof of part (i)

(ii) Consider the leader's decisions facing a null recommendation. All in all we have 6 ICs to consider:

$$\begin{aligned}&(i, ii)\ \ (\{N, \neg N\}, \alpha) \times (\hat{\emptyset}) \to a_1, &&(iii)\ \ (\neg N, \emptyset) \times (\hat{\emptyset}) \to a_1,\\ &(iv, v)\ \ (\{N, \neg N\}, \beta) \times (\hat{\emptyset}) \to b_1, &&(vi)\ \ (N, \emptyset) \times (\hat{\emptyset}) \to b_1.\end{aligned}$$

But among these, in case (iv) $(N, \beta) \times (\hat{\emptyset}) \to b_1$, the leader will not deviate because implementing b_1 is the ex-ante efficient decision that also helps her retaining leadership in period 2; in case (vi) $(N, \emptyset) \times (\hat{\emptyset}) \to b_1$, again the leader has no incentive to deviate because equal priors $\Pr(a_1) = \Pr(b_1) = 1/2$ imply choice of b_1 is as efficient ex ante as a_1, and the proposed implementation helps the leader to retain leadership. So we need to consider the IC's only for the following four cases:

$$\text{(i)}\ \ (\mathsf{N}, \alpha) \times (\widehat{\emptyset}) \to a_1, \qquad \text{(ii)}\ \ (\neg\mathsf{N}, \alpha) \times (\widehat{\emptyset}) \to a_1,$$
$$\text{(iii)}\ \ (\neg\mathsf{N}, \emptyset) \times (\widehat{\emptyset}) \to a_1, \qquad (\nu)\ \ (\neg\mathsf{N}, \beta) \times (\widehat{\emptyset}) \to b_1,$$

which we next analyze in the stated order.

IC for case (i)**.** The $(\mathsf{N}, \alpha, \xi)$ leader's payoff from implementing a_1 equals

$$\xi(sV_1) + \big[\nu_2\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\big\}sV_2 + (1-\nu_2)\frac{sV_2}{2}\big],$$

whereas her payoff from implementing b_1 equals

$$(1-\xi)(sV_1) + \big[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\big](sV_2).$$

The leader would implement a_1 rather than deviate to b_1 if

$$(2\xi - 1)(sV_1) + (1-\nu_2)\Bigg[-\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\big\} + \frac{1}{2}\Bigg]sV_2 \geq 0. \tag{B.8}$$

IC for case (ii)**.** The $(\neg\mathsf{N}, \alpha, \xi)$ leader's payoff from implementing a_1 equals

$$\zeta + \xi(sV_1) + \big[\nu_2\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\big\} + (1-\nu_2)\frac{1}{2}\big]sV_2,$$

whereas her payoff from implementing b_1 equals

$$(1-\xi)(sV_1) + \big[\zeta + \frac{1}{2}(sV_2)\big].$$

The leader would implement a_1 rather than deviate to b_1 if

$$(2\xi - 1)(sV_1) + \nu_2\Bigg[\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\big\} - \frac{1}{2}\Bigg]sV_2 \geq 0. \tag{B.9}$$

IC for case (iii)**.** The $(\neg\mathsf{N}, \emptyset, \xi)$ leader's payoff from implementing a_1 equals

$$\zeta + \frac{1}{2}(sV_1) + \Bigg[\nu_2\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\big\} + (1-\nu_2)\frac{1}{2}\Bigg]sV_2,$$

whereas her payoff from implementing b_1 equals

$$\frac{1}{2}(sV_1) + \big[\zeta + \frac{1}{2}(sV_2)\big].$$

The leader would implement a_1 rather than deviate to b_1 if

$$\nu_2\Bigg[\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\big\} - \frac{1}{2}\Bigg]sV_2 \geq 0. \tag{B.10}$$

IC for case (ν)**.** The $(\neg\mathsf{N}, \beta, \xi)$ leader's payoff from implementing a_1 equals

$$\zeta + (1-\xi)(sV_1) + \big[\nu_2\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\big\} + (1-\nu_2)\frac{1}{2}\big]sV_2,$$

whereas her payoff from implementing b_1 equals

$$\xi(sV_1) + \big[\zeta + \frac{1}{2}(sV_2)\big].$$

The leader would implement a_1 rather than deviate to b_1 if

$$(1-\xi)(sV_1) + \nu_2\Big[\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\} - \frac{1}{2}\Big]sV_2 \geq 0. \tag{B.11}$$

Note that condition (B.10) guarantees conditions (B.9) and (B.11). Thus when the leader faces a null recommendation, conditions (B.8) and (B.10) will ensure the leader's incentive compatibility.

Conditions (B.7), (B.8) and (B.10) are collated below for ease of reference:

$$\frac{\lambda-\xi}{(1-\xi)\lambda+\xi(1-\lambda)}(sV_1) - \nu_2\Big[\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\} - \frac{1}{2}\Big]sV_2 \geq 0;$$

$$(2\xi-1)(sV_1) + (1-\nu_2)\Big[-\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\} + \frac{1}{2}\Big]sV_2 \geq 0;$$

$$\nu_2\Big[\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\} - \frac{1}{2}\Big]sV_2 \geq 0.$$

This completes the proof of part (ii). **Q.E.D.**

***Proposition 12* (Truthful advice)** *Let agent 1 be the period 1 leader who is of talent ξ (i.e., of lesser talent), and fix the principal's and the leader's strategies as specified in $\mathcal{E}_1$. Under the incentive compatibility conditions* (B.14) $\geq$ max{(B.15), (B.16)} *(derived in the proof), the equilibrium of the Stage T_{12} game of $\mathcal{E}_1$ will have the following property:* A talented deputy *(agent 2) has no incentive to misreport her signal.*

Propositions 11 and 12, together, make up the informative advice equilibrium. Under truthful advice project implementations in period 1 is ex-ante efficient, as a corrupt leader without a clear recommendation from the deputy and when her own information is null leans towards the project offering bribes when either project is equally efficient. In all other situations, a corrupt leader does no different from an unbiased leader, and the latter behaves honestly. Given that period 2 equilibrium is sincere, equilibrium $\mathcal{E}_1$ is thus the best the principal can hope to achieve. This equilibrium, however, need not always exist (see Fig. 14). It exists if the experts observe an informative signal with a high enough probability (ϵ small), the leader who is less talented is unbiased with a high enough probability ($\nu_1 > \nu_1''$) and the deputy's honesty is neither too low nor too high ($\nu_2 \in (\underline{\nu}, \overline{\nu})$), as stated in Proposition 9, later proved in this Appendix, and illustrated in Fig. 14.

Proof of Proposition 12. We start with an ancillary result.

Step 1. Fix any signal-talent pair of the deputy in period 1, $t_d \equiv (\sigma_d, \tau_d) \in T_d \setminus \{N, \neg N\}$. Let the deputy's period 1 equilibrium report, as in $\mathcal{E}_1$, consisting of her signal-talent pair or a null report be ρ_d, and denote by $\hat{\rho}_d$ any feasible report. Then for the strategies and beliefs as earlier specified in $\mathcal{E}_1$, the following will be true:

(i) A biased t_d-deputy's payoff from reporting any $\hat{\rho}_d$ exceeds an unbiased t_d-deputy's payoff from reporting the same $\hat{\rho}_d$.

(ii) Suppose that for the same parametric restriction $\mathcal{P}$ on model primitives (V_1, V_2, ϵ, λ, ξ, ζ, ν_1, ν_2, q, etc.), an unbiased t_d-deputy's payoff from equilibrium reporting ρ_d (weakly) exceeds a biased t_d-deputy's payoff from reporting some other $\hat{\rho}_d$, $\hat{\rho}_d \neq \rho_d$. Then given the restriction $\mathcal{P}$, no t_d-deputy, whether biased or unbiased, deviates to $\hat{\rho}_d$.

Part (i) is clear: period 1 payoff of a deputy is independent of whether she is biased or not, whereas by Assumption 2(a), period 2 payoff for the biased deputy will exceed that of the unbiased deputy in case regime switch happens with a positive probability (period 2 payoffs will be equal otherwise).

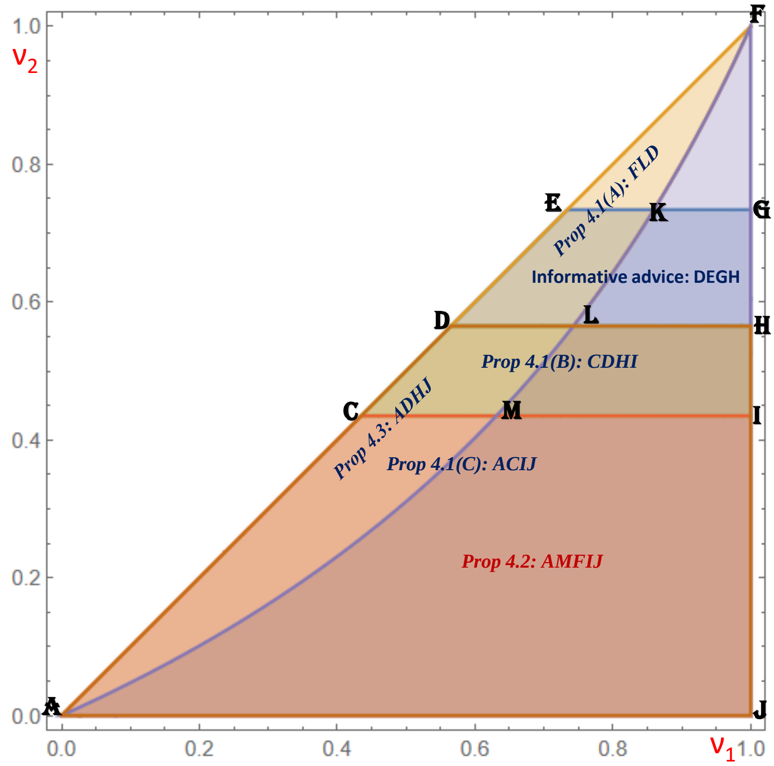


Figure 14: Informative advice and Yesman-I equilibria, fixing $\xi = 0.65, \lambda = 0.85, \epsilon = 0.1, s = 0.1, V_1 = 2000, V_2 = 4200, \zeta \geq 350$.

To establish part (ii), suppose, first, the t_d-deputy is biased. Then,

$$\begin{aligned}
&\text{biased } t_d\text{-deputy's payoff from equilibrium reporting } \rho_d \\
&\geq \text{an unbiased } t_d\text{-deputy's payoff from equilibrium reporting } \rho_d \\
&\geq \text{biased } t_d\text{-deputy's payoff from any other reporting } \hat{\rho}_d,
\end{aligned} \tag{B.12}$$

where the first inequality follows from part (i), and the second inequality is true by hypothesis in part (ii). Thus a biased t_d-deputy will not deviate.

Next suppose the t_d-deputy is unbiased. Then,

$$\begin{aligned}
&\text{an unbiased } t_d\text{-deputy's payoff from equilibrium reporting } \rho_d \\
&\geq \text{biased } t_d\text{-deputy's payoff from reporting } \hat{\rho}_d \\
&\geq \text{an unbiased } t_d\text{-deputy's payoff from reporting } \hat{\rho}_d,
\end{aligned} \tag{B.13}$$

where the first inequality is true by hypothesis, and the second inequality follows from part (i). Thus an unbiased t_d-deputy will not deviate. $\|$

Consider the deputy's strategy in $\mathcal{E}_1$. Given the preceding result, in particular part (ii), it is sufficient to establish that an unbiased t_d-deputy's equilibrium payoff from ρ_d exceeds a biased t_d-deputy's deviation payoff from $\hat{\rho}_d$, where $t_d \equiv (\sigma_d, \tau_d) \in T_d \setminus \{N, \neg N\}$. In what follows, it is this sufficient condition that we shall always check for. This reduces the number of conditions to be checked for by half.

Step 2. Consider the deviation incentives of the deputy who, by assumption, is talented (λ).

(A) [Deputy has signal β.] Consider an unbiased deputy's equilibrium payoff from reporting $\hat{\beta}$:

$$\lambda\, sV_1 + \big[\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big]sV_2. \tag{B.14}$$

We next consider the deviation payoffs.

• *Deviation to* $\hat{\alpha}$. Reporting truthfully dominates a deviation to $\hat{\alpha}$ since such a deviation does not lead to any regime switch yet period 1 payoff will be lower as a_1 will be implemented, which is inefficient. Thus it is sufficient to consider deviation to $\hat{\emptyset}$.

• *Deviation to* $\hat{\emptyset}$. It is sufficient to consider the deviation payoff of a biased deputy. This is because the first period payoffs are the same, but a biased deputy has a higher second period payoff since reporting $\hat{\emptyset}$ can lead to regime switching.

The payoff of $(\neg N, \beta)$ deputy from deviation to $\hat{\emptyset}$ reporting is:

$$\begin{aligned}
&(1-\epsilon)\Bigg[\Pr(\omega_1 = a_1|\sigma_2 = \beta)\Bigg\{\Pr(\sigma_1 = \alpha|\omega_1 = a_1)\big[sV_1 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] \\
&\qquad + \Pr(\sigma_1 = \beta|\omega_1 = a_1)\big[0 + \big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}sV_2\big]\Bigg\} \\
&\quad + \Pr(\omega_1 = b_1|\sigma_2 = \beta)\Bigg\{\Pr(\sigma_1 = \alpha|\omega_1 = b_1)\big[0 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] \\
&\qquad + \Pr(\sigma_1 = \beta|\omega_1 = b_1)\big[sV_1 + \big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}sV_2\big]\Bigg\}\Bigg] \\
&+ \epsilon(1-\nu_1)\Bigg[\Pr(\omega_1 = a_1|\sigma_2 = \beta)\big[sV_1 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] + \Pr(\omega_1 = b_1|\sigma_2 = \beta)\big[0 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big]\Bigg] \\
&+ \epsilon\nu_1\Bigg[\Pr(\omega_1 = a_1|\sigma_2 = \beta)\big[0 + \big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big)sV_2\big] \\
&\qquad + \Pr(\omega_1 = b_1|\sigma_2 = \beta)\big[sV_1 + \big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big\}sV_2\big]\Bigg] \\
&= (1-\epsilon)\Bigg[(1-\lambda)\Bigg\{\xi\big[sV_1 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] + (1-\xi)\big[\big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}sV_2\big]\Bigg\} \\
&\qquad + \lambda\Bigg\{(1-\xi)\big[\zeta + \frac{1}{2}(sV_2)\big] + \xi\big[sV_1 + \big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}sV_2\big]\Bigg\}\Bigg] \\
&\quad + \epsilon(1-\nu_1)\Bigg[(1-\lambda)\big[sV_1 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] + \lambda\big[\zeta + \frac{1}{2}(sV_2)\big]\Bigg] \\
&\quad + \epsilon\nu_1\Bigg[(1-\lambda)\big[\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big)sV_2\big] + \lambda\big[sV_1 + \big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big\}sV_2\big]\Bigg] \\
&= \Bigg[(1-\epsilon)\xi + \epsilon\big\{(1-\nu_1)(1-\lambda) + \nu_1\lambda\big\}\Bigg]\Bigg[sV_1\Bigg] + \Bigg[(1-\epsilon)\big\{(1-\lambda)\xi + \lambda(1-\xi)\big\} + \epsilon(1-\nu_1)\Bigg]\Bigg[\zeta + \frac{1}{2}(sV_2)\Bigg] \\
&\quad + \Bigg[(1-\epsilon)\big\{(1-\lambda)(1-\xi) + \lambda\xi\big\}\big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\} \\
&\qquad + \epsilon\nu_1\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big\}\Bigg]\Bigg[sV_2\Bigg].
\end{aligned} \tag{B.15}$$

So, the IC condition of a λ-talent deputy with signal β is:

$$
\begin{aligned}
&(B.14) - (B.15)\\
&= \lambda\, sV_1 + \Big[\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\Big]sV_2\\
&\quad - \Big[(1-\epsilon)\xi + \epsilon\big\{(1-\nu_1)(1-\lambda) + \nu_1\lambda\big\}\Big]\Big[sV_1\Big]\\
&\quad - \Big[(1-\epsilon)\big\{(1-\lambda)\xi + \lambda(1-\xi)\big\} + \epsilon(1-\nu_1)\Big]\Big[\zeta + \frac{1}{2}(sV_2)\Big]\\
&\quad - \Big[(1-\epsilon)\big\{(1-\lambda)(1-\xi) + \lambda\xi\big\}\big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}\\
&\qquad + \epsilon\nu_1\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big\}\Big]\Big[sV_2\Big] \geq 0.
\end{aligned}
$$

(B) [Deputy has signal α.] From Step 1 earlier, it suffices to consider an unbiased α-signal deputy's payoff from reporting $\hat{\alpha}$. Recalling that the leader (agent 1) implements a_1 and that there is no regime switch the payoff is:

$$
\lambda\, sV_1 + \Big[\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\Big]sV_2,
$$

which is same as (B.14).

• *Deviation to* $\hat{\emptyset}$. It is sufficient to consider deviation by a biased deputy to reporting $\hat{\emptyset}$, when she has a payoff of:

$$
\begin{aligned}
&(1-\epsilon)\Big[\Pr(\omega_1 = a_1|\sigma_2 = \alpha)\Big\{\Pr(\sigma_1 = \alpha|\omega_1 = a_1)\big[sV_1 + \{\zeta + \frac{1}{2}(sV_2)\}\big]\\
&\qquad + \Pr(\sigma_1 = \beta|\omega_1 = a_1)\big[0 + \big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}sV_2\big]\Big\}\\
&\quad + \Pr(\omega_1 = b_1|\sigma_2 = \alpha)\Big\{\Pr(\sigma_1 = \alpha|\omega_1 = b_1)\big[0 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big]\\
&\qquad + \Pr(\sigma_1 = \beta|\omega_1 = b_1)\big[sV_1 + \big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}sV_2\big]\Big\}\Big]\\
&+ \epsilon(1-\nu_1)\Big[\Pr(\omega_1 = a_1|\sigma_2 = \alpha)\big[sV_1 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] + \Pr(\omega_1 = b_1|\sigma_2 = \alpha)\big[0 + \big\{\zeta + \frac{1}{2}(sV_2)\}\big]\Big]\\
&+ \epsilon\nu_1\Big[\Pr(\omega_1 = a_1|\sigma_2 = \alpha)\big[0 + \big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big)sV_2\big]\\
&\quad + \Pr(\omega_1 = b_1|\sigma_2 = \alpha)\big[sV_1 + \big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big\}sV_2\big]\Big]\\
&= (1-\epsilon)\Big[\lambda\Big\{\xi\big[sV_1 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] + (1-\xi)\big[\big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}sV_2\big]\Big\}\\
&\qquad + (1-\lambda)\Big\{(1-\xi)\big[\big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] + \xi\big[sV_1 + \big\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big\}sV_2\big]\Big\}\Big]\\
&\quad + \epsilon(1-\nu_1)\Big[\lambda\big[sV_1 + \big\{\zeta + \frac{1}{2}(sV_2)\big\}\big] + (1-\lambda)\big[\zeta + \frac{1}{2}(sV_2)\big]\Big]\\
&\quad + \epsilon\nu_1\Big[\lambda\big[\big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big\}sV_2\big] + (1-\lambda)\big[sV_1 + \big\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big\}sV_2\big]\Big]
\end{aligned}
$$

$$= \Big[(1-\epsilon)\xi + \epsilon(1-\nu_1)\lambda\Big]\Big[sV_1\Big]$$
$$+ \Big[(1-\epsilon)\{\lambda\xi + (1-\lambda)(1-\xi)\} + \epsilon(1-\nu_1)\Big]\Big[\zeta + \frac{1}{2}(sV_2)\Big]$$
$$+ \Big[(1-\epsilon)\{\lambda(1-\xi) + (1-\lambda)\xi\}\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\}$$
$$+ \epsilon\nu_1\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\}\Big]\Big[sV_2\Big]. \tag{B.16}$$

- *Deviation to* $\hat{\beta}$. The payoff to the biased deputy from this deviation report is:

$$(1-\lambda)sV_1 + \big[\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\big]sV_2. \tag{B.17}$$

The condition (B.14)≥(B.17) is always satisfied. So, the IC condition of an (α, λ)-deputy is:

$$(B.14) - (B.16)$$
$$= \lambda\, sV_1 + \Big[\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\Big]sV_2$$
$$- \Big[(1-\epsilon)\{\lambda\xi + (1-\lambda)(1-\xi)\} + \epsilon(1-\nu_1)\Big]\Big[\zeta + \frac{1}{2}(sV_2)\Big]$$
$$- \Big[(1-\epsilon)\{\lambda(1-\xi) + (1-\lambda)\xi\}\{\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big) + (1-\nu_1)\frac{1}{2}\}$$
$$+ \epsilon\nu_1\big((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2\frac{1}{2}\big)\Big]\Big[sV_2\Big] \geq 0.$$

(C) [Deputy has signal $\emptyset$.] Deviation to a report of $\hat{\alpha}$ implies (i) regime switch does not happen whereas with a report of $\hat{\emptyset}$ regime switch happens with a positive probability, (ii) ex ante, project choice becomes worse when the leader observes signal β, and no better or worse when she observes α or $\emptyset$, the reason for the latter case (of $\sigma_\ell = \emptyset$) being that the two states are equally likely a priori. So the deputy should not deviate to reporting $\hat{\alpha}$.

A symmetric argument applies to rule out deviation to $\hat{\beta}$ report. **Q.E.D.**

Proof of Proposition 9. Step 1. We first consider the deputy's incentive compatibility conditions. From continuity it is sufficient to show that these hold, for $\epsilon = 0$ and $\nu_1 = 1$. Note that

$$(B.14)|_{\epsilon=0,\nu_1=1} = \lambda\, sV_1 + \lambda\, sV_2, \tag{B.18}$$
$$(B.15)|_{\epsilon=0,\nu_1=1} = \xi\, sV_1 + \{(1-\lambda)\xi + (1-\xi)\lambda\}(\zeta + \frac{sV_2}{2}) + \{(1-\lambda)(1-\xi) + \lambda\xi\}\lambda sV_2, \tag{B.19}$$
$$(B.16)|_{\epsilon=0,\nu_1=1} = \xi\, sV_1 + \{(1-\lambda)(1-\xi) + \lambda\xi\}(\zeta + \frac{sV_2}{2}) + \{(1-\lambda)\xi + (1-\xi)\lambda\}\lambda sV_2. \tag{B.20}$$

Note that $[(B.14) - (B.15)]|_{\epsilon=0,\nu_1=1} \geq 0$, if and only if

$$(\lambda-\xi)sV_2 \geq sV_2[\{(1-\lambda)\xi + (1-\xi)\lambda\}\frac{1}{2} + \{(1-\lambda)(1-\xi) + \lambda\xi\}\lambda - \lambda] + \{(1-\lambda)\xi + (1-\xi)\lambda\}\zeta.$$

Next, $[\{(1-\lambda)\xi + (1-\xi)\lambda\}\frac{1}{2} + \{(1-\lambda)(1-\xi) + \lambda\xi\}\lambda - \lambda] < 0$, since $(1-\lambda)\xi + (1-\xi)\lambda\}\frac{1}{2} + \{(1-\lambda)(1-\xi) + \lambda\xi\}\lambda$ is a convex combination of λ and $1/2$, and hence less than λ. The inequality thus holds since $(\lambda-\xi)sV_2 \geq \{(1-\lambda)\xi + (1-\xi)\lambda\}\zeta$.

Next note that $[(B.14) - (B.16)]|_{\epsilon=0,\nu_1=1} \geq 0$, if and only if

$$(\lambda-\xi)sV_2 \geq sV_2[\{(1-\lambda)\xi + (1-\xi)\lambda\}\lambda + \{(1-\lambda)(1-\xi) + \lambda\xi\}\frac{1}{2} - \lambda] + \{(1-\lambda)(1-\xi) + \xi\lambda\}\zeta.$$

Next, $[\{(1-\lambda)\xi + (1-\xi)\lambda\}\lambda + \{(1-\lambda)(1-\xi) + \lambda\xi\}\frac{1}{2} - \lambda] < 0$, since $[\{(1-\lambda)\xi + (1-\xi)\lambda\}\lambda + \{(1-\lambda)(1-\xi) + \lambda\xi\}\frac{1}{2}]$ is a convex combination of λ and $1/2$, and hence less than λ. The inequality thus holds since $(\lambda - \xi)sV_2 \geq \{(1-\lambda)(1-\xi) + \xi\lambda\}\zeta$.

Step 2. We then consider the leader's IC conditions. Given that $2\overline{V_2} > V_2$, both (B.7) and (B.8) are satisfied for $\nu_2 = \frac{1}{2}$. Finally, (B.10) is always satisfied since $\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \epsilon^2(\frac{1}{2})\}$ is a convex combination of λ, ξ and $1/2$, and hence exceeds $1/2$. **Q.E.D.**

Proof of Proposition 10. Recall that the principal's surplus under informative advice equilibrium equals[40]

$$\begin{aligned}
(17) = &\left((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\right)(s_P V_1) \\
&+ \left[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\frac{\xi}{4}\right]\left(\nu_1\left\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\right\} + (1-\nu_1)\frac{1}{2}\right)(s_P V_2) \\
&+ \left[\frac{\epsilon}{2} + \frac{\epsilon^2}{2}\right]\left(\nu_2\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\} + (1-\nu_2)\frac{1}{2}\right)(s_P V_2) \\
&+ \epsilon^2\left((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\right)(s_P V_2).
\end{aligned}$$

1. Thus, the difference in the principal's surpluses, (17) − (6), can be written as follows after some simplifications:

$$\begin{aligned}
= &\left((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2} - \nu_1(1-\epsilon)(\xi - \frac{1}{2}) - \frac{1}{2}\right)(s_P V_1) \\
&+ \left[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\frac{\xi}{4}\right]\left(\nu_1\left\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\right\} + (1-\nu_1)\frac{1}{2}\right)(s_P V_2) \\
&+ \left[\frac{\epsilon}{2} + \frac{\epsilon^2}{2} - \nu_1\frac{1}{2}(1-\epsilon) - (1-\nu_1)\right]\left(\nu_2\left\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\right\} + (1-\nu_2)\frac{1}{2}\right)(s_P V_2) \\
&+ \left[\epsilon^2 - \frac{\nu_1}{2}(1+\epsilon)\right]\left((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\right)(s_P V_2).
\end{aligned}$$

(i) The first panel of Fig. 12 shows that there exist parameter values where (17) − (6)< 0, i.e. the surplus under the informative advice equilibrium is less than that under the Yesman equilibrium.

(ii) The second panel of Fig. 12 shows that there exist parameter values where (17) − (6)> 0, i.e. the surplus under the informative advice equilibrium is greater than that under the Yesman equilibrium.

2. The difference in the principal's surpluses, (17) − (9), can be written as follows after some simplifications:

$$\begin{aligned}
= &\left((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2} - \nu_1(1-\epsilon)(\xi - \frac{1}{2}) - \frac{1}{2}\right)(s_P V_1) \\
&+ \left[(1-\epsilon)\lambda + \epsilon(1-\epsilon)\frac{\xi}{4} - 1\right]\left(\nu_1\left\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\right\} + (1-\nu_1)\frac{1}{2}\right)(s_P V_2) \\
&+ \left[\frac{\epsilon}{2} + \frac{\epsilon^2}{2}\right]\left(\nu_2\{(1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\} + (1-\nu_2)\frac{1}{2}\right)(s_P V_2) \\
&+ \epsilon^2\left((1-\epsilon)\lambda + \epsilon(1-\epsilon)\xi + \frac{\epsilon^2}{2}\right)(s_P V_2).
\end{aligned}$$

[40] The complete derivation is available on request

(i) The first panel of Fig. 13 shows that there exist parameter values such that (17) − (9)< 0, i.e. the surplus under the informative advice equilibrium is less than that under the Yesman equilibrium.

(ii) The second panel of Fig. 13 shows that there exist parameter values such that (17) − (9)> 0, i.e. the surplus under the informative advice equilibrium is greater than that under the Yesman equilibrium.