Delegation without Commitment

Scott Baker
Washington University in St. Louis Law School

Lewis A. Kornhauser
New York University Law School

17 April 2017

Abstract

We study a model of delegation without commitment. The principal must resolve a sequence of claims about which she has incomplete information. She can delegate to an agent with superior information but she cannot commit to respecting the agent's decision. We show that, even in the absence of commitment, it is rational for the principal to defer to the agent in a broad set of instances. In equilibrium, that is, it is in the principal's interest to defer to the agent's decision.

Our model is set in a two-dimensional space of claims. One dimension reflects "global" facts that are known to everyone; the other dimension reflects "local" facts that are known only to the agent. The principal has a preferred partition of the space of claims into two sets: those that should be decided "valid" and those that should decided "not valid."

The agent may, with some probability, be biased in the sense that it thinks a different partition is best. A one-period model identifies the basic tradeoff the principal faces between granting discretion to a potentially biased agent and deciding for herself. Deference risks that a biased agent will decide the case wrongly while ruling on the basis of global facts only ignores local facts and also risks error. The degree of deference granted balances these two risks. Intuitively, the model shows that cases where the global facts are inconclusive ("hard cases") are the ones where the grant of discretion is most valuable (i.e. "make good law".)

We then extend the model to two periods and investigate the relationship between principal and agent and the extent of delegation.We show that, in period 1, the principal increases the region of delegation to in-

crease the likelihood of learning the agent's type. In response, the biased agent in period 1 mimics, over a range of values of global facts, the conscientious agent. This mimicry unambiguously improves the principal's expected period 1 payoff. In period 2, the principal, following a period in which the agent's type remains concealed, expands the region of delegation or, following revelation of the agent's type, dramatically restricts it.

The model applies to a wide variety of judicial and bureaucratic settings, both public and private, in which claims are resolved.

1 Introduction

The principal-agent relation sits at the heart of every organization. How can a principal induce an agent to act in the principal's interests when the agent has both superior information to and divergent preferences from the principal? Following Holmstrom's pioneering work (Holmstrom 1977, 1984), economists and political scientists have developed a rich and fruitful formal theory.

This formal theory has typically been set in an economic setting in which the principal has a broad array of mechanisms through which she can control her agent, most prominently dismissal, bonuses, and penalties. In many political settings, however, these standard control mechanisms are muted at best and often non-existent. The federal judiciary presents the starkest contrast to the standard economic setting. Federal judges in the United States are appointed for life, their salaries cannot be changed on the basis of performance, and they have few if any prospects of promotion. In these circumstances, the principal has few if any mechanisms to control her agent.

Federal judges present an extreme example of the limited range of control mechanisms available to control public officials. But, in many jurisdictions, most public officlals are protected by civil service regulations that limits the principal's ability to dismiss an agent, even for cause, and limit the ability of the principal to reward outstanding performance or to punish unacceptable performance through the agent's pay. We present a formal model of what are essentially reputation mechanisms of control.

We model a setting that is typical of many government and private organizations that handles claims. Some third party has a claim against the organization and the organization must determine its validity. The agent investigates the claim and either grants or denies it. Its investigation develops both "local"

and "global" information. The agent's decision is reviewed by the principal, who can either affirm or reverse. In conducting its review, the principal has access to the global information – the text of the contract in dispute, say. It does not have access to the local information – the credibility of the star witness.

The sole means through which the principal can control the agent is through review of the agent's decision. This framework captures a phenomenon common to many dispute resolution settings. Most obviously, it is a bare model of adjudication in which trial courts find facts and resolve disputes subject to review by an appellate court. It also captures judicial review of rulemaking by administrative agencies.[1] Many government bureaucracies, however, engage in similar activities. The Social Security Adminsitration, the Center for Medicare and Medicaid Services and the Department of Citizen and Immigration Services all resolve millions of claims a year concerning disability, health care provision, asylum and deportation. Insurance companies have similar departments as do retailers.

Our model has several distinctive features. First, as noted, the information structure contains both global information and local private information. Unlike many prior models, the principal isn't ignorant; he is partially informed. Second, the global information is more valuable in some cases than others. Some court cases, for example, can be adjudicated on the text of the contract (text easily observed by the appellate court). Other cases require consideration of both the text and testimony about the parties intentions at the time of formation (the latter facts observable only to the trial court). Our model explains what cases are apt to fall in this bin.

Third, the principal is unable to commit to a review strategy. She cannot commit to grant authority to the agent (through, say, the sale of the means of production). Finally, the model assumes that the agent has a concern for reputation. This concern takes two forms in the model. We assume that the agent incurs a cost from reversal. One might understand this cost as the agent's concern for her reputation or as a form of self-respect. But we also study a more consequential form of reputation by analyzing a dynamic game in which the agent has the opportunity to build a reputation in the first period that increases its freedom of action in the second period. To do this, we structure the game as one of incomplete information in which the agent may be either

[1] This setting is somewhat more complex than the adjudication of a common law claim as one might argue that Congress rather than the appellate court is the true principal of the administrative agency.

"conscientious" or "biased". A conscientious agent decides non-strategically in conformity with the preferences of the principal. A biased agent, by contrast, acts strategically to further its own preferences that diverge from those of the principal.

The baseline model is a one-period, static model. We identify the principal's optimal strategy as one of "delegation" in which, for a broad range of global facts, the principal affirms the agent's decision with probability one. This policy allows the principal to exploit the agent's superior knowledge of local facts. If the agent were conscientious with probability 1, the principal would fully delegate its decision ot the agent. Full delegation to a biased agent, however, has costs as the biased agent will wrongly decide many cases. When global facts, however, are "extreme" in the sense that the divergence between the principal's and the agent's preferences are sufficiently great, the principal controls the agent's behavior with a strategy that reverses with positive probability. Notably, the principal only reverses decisions that are "unexpected" given the location of the global facts. The principal has a sense of what the result should be by looking at the global fact. Red flays arise – and with them a probability of reversal – when the decision differs from what is expected. By contrast, expected or routine decisions are affirmed. Finally, the greater the divergence between the principal's and the agent's preferences, the greater the likelihood of reversal of an unexpected decision.

We then turn to a two-period game. We study the incentives on the agent to build reputation and how the extent of deference by the principal may vary over time. The model offers three major insights. The agent, in period 1, rules against its own preferences today to appear like a more conscientious agent as conscientious agents have discretion over a broader class of cases in period 2. The principal thus has an incentive in period 1 to widen the range of discretion in order to learn the agent's type. Concomitantly, the principal will use whatever it learned in period 1 to either broaden or narrow the range of discretion. Second, the reputational mechanism is not sufficient to control the agent perfectly. Third, not all cases test "loyalty" the same way. We show that the reaction – in terms of reputation building or destruction – is not same across all cases. For many cases, resolution consistent with the principal's preferences is expected, and as a result, the agent doesn't earn esteem from unproblematic resolutions of the case. For other cases, a consistency between the preferences of principal and agent is less likely and the prospects for revelation of the agent's type are correspondingly greater.

The literature in both economics and political science on the principal-agent relation is vast. The formal study of this problem began with Holmstrom (1977, 1984). In the basic framework, the principal has a project in the completion of which she may wish to enlist the services or knowledge of an agent whose preferences differ from hers. The principal has imperfect information about the state of the world. The principal observes the agent's decision but is unable to condition the agent's incentives on both her decison and the state of the world. The principal must decide when to let the agent take some action on the real line.

This literature has developed in two main directions. The first, *delegation* strand, follows Holmstrom's assumption that contracts between the the principal and agent are fully enforceable. Specifically, the principal can commit to uphold the agent's decision if her action falls within a certain range. The agent is granted authority over, say, the use of a critical input in production. Economists have studied how a principal might delegate decision-making authority to a biased agent; Athey et al. 2005). The delegation models in political sceince have a similar structure (Huban & Shipan 2006). The incentive compatibility mechanism often is an interval delegation (Amador & Bagwell 2014). The principal allows the agent to take actions within the interval and prohibits those outside the interval. Such models do not readily translate into the context we study. Bureaucrats, both public and private often make decisions that are binary; this binary relation is perhaps clearest in the context of adjudication in which the trial court (as agent) and the appellate courts (as principal) face such question as: Should the defendant be liable or not? Should the expert testimony be admitted or not? Does the constitution allow corporations to lodge religious objections or not? But health insurers, for instance, face a similar strategic structure when asked to determine whether a given policy covers a particular treatment. In other words, the action space is dichotomous; the principal thus cannot create an interval. In our model, however, the principal has partial knowledge of the state of the world; it knows some "global" facts but not local ones. By setting the model in two dimensions, we are able to translate the action interval into an "interval" in the state space.

The second, *cheap talk* strand grows from Crawford and Sobel (1982). As before, the principal has to select a project; her optimal choice depends on the state of the world which she does not know with certainty. The agent has better information about the state of the world than the principal. Crawford and Sobel study the conditions under which the agent can, through "cheap talk," influence

the principal's decision. Talk is "cheap" because, in these models, whatever is said has no effect on the payoff of the principal or agent. In our model, the agent's message space is its ruling "valid" or"not valid". Its ruling may convey information to the principal about its local information available only to it.

Our model may be interpreted as either a delegation or a cheap talk model depending on the context. Essentially, in those circumstances in which the principal's reversal of the agent's action imposes an outcome, our model looks like cheap talk. But, in some circumstances, particularly in the review of administrative agency rule-making, when an appellate court reverses the decision below it vetoes the policy chosen below. In this context, our model seems more like a delegation model.[2] More specifically, though there is no *external* commitment mechanism (as seems to be assumed in many political science models of delegation), the principal has a self-enforcing commitment to broad delegation that characterizes the equilibrium.

Our model draws on literature on reputation in games with incomplete information to integrate decisions about the scope of deference with a dynamic model of reputation-building. It is well-known that agents might act differently in the face of reputational consideration (Benabou & Larique 1992; Morris 2001). Biased agents might mimic conscientious ones; conscientious ones might manipulate reported information to avoid appearing biased (Morris 2001). Our model falls within the class of reputation models where decisions do not perfectly reveal the agent's type. For any case within the bounds of discretion, both conscientious and biased agents might find the claim valid or not valid. The information relevant to reputation is the "rate" at which they do so. Different decision rates allow the principal to learn, albeit imperfectly, about the agent's preferences. Our model directly addresses why agents want the principals to hold them in high esteem. There are no monetary transfers between principals and agents. In our model, the incentive for reputation building all comes from the agent's desire for future discretion.

Finally, legal scholars have spent time on the standard of review, the key deference step in a judicial opinion. Administrative law scholars have explored the positive question of whether appellate courts treat decisions by agencies different from decisions by trial courts and the normative question of whether they should.[3]

[2] Bendor et al (2001) and Gailmard and Patty (2012) offer reviews of the principal-agent literature from a political science perspective. Gailmard and Patty briefly discuss a modified delegation model with a veto.

[3] Our model contributes to the literature on judicial hierarchy as well. Reinganum and

The paper proceeds as follows. Section 2 presents the one-period model and derives the optimal bounds of discretion in the static setting. Section 3 considers a two period model. It explores how deference responds to trial court or agency reputation. It also pinpoints which cases present the most potential to build or destroy the reputational capital of trial courts and agencies. Section 5 concludes.

2 The Model

2.1 Preliminaries

A principal and an agent interact over two periods. (Throughout we refer to the principal as "she" and the agent as "he".) In each period, a "claim" s_t randomly arrives for resolution. The agent resolves the claim. The claim is appealed and reviewed by the principal. A claim $s_t = (x_t,y_t)$ consists of two facts, x_t and y_t, each of which is (independently) drawn from a uniform distribution over the unit interval $[0, 1]$. The global fact, x_t, is observable to both the agent and the principal. The local fact, y_t, is observable to the agent only. The space S of possible claims is thus the unit square.[4] As an example, consider judicial dispositions in a hierarchy. The principal is the appellate court. The agent is the trial court. A global fact might be the text of the contract under dispute. This text is easily observed on appeal. The local fact might the demeanor and credibility of the star witness at trial, a variable which is not observed on appeal.

The timing of the stage game in period t is

1. A claim x_t,y_t arises.

2. The agent observes the claim and announces a decision $d_t(x_t, y_t)$. The decision declares the claim "valid" ($d_t = v$) or "not valid." ($d_t = nv$)

Daughety (2000) model trial court and appellate court interactions. The trial court finds facts and applies the law. This law isn't known with certainty, however. Indeed, the appellate court has more information about the law than the trial court. The appellate court takes the findings of fact as given and then makes inferences about the law based on his own informative signal and the decision to appeal. The trial court and the appellate court share preferences, but have different information about what the ultimate Supreme Court prefers. Our model presents a conflict between courts in the hierarchy over the appropriate dispositions and allows the trial court or agency to improve or destroy its reputation over time. Lax (2012) provides a model of rules and standards, where a standard provide the trial court with more flexibility, but is also easier for the appellate court to specify. Many of the classic tradeoffs arises between rules and standards. Kastellec (2016) surveys the political science literature on judicial hierarchy.

[4] In formal models of courts, S is typically called the "case space" and our model deploys the formalism developed there. See Kornhauser [1992] and Lax [2011].

3. The principal observes the agent's decision and the global fact. She then either reverses ($\gamma_t(x_t, d_t) = 1$) or affirms ($\gamma_t(x_t, d_t) = 0$).

4. Payoffs $u^A(s_t, d_t^F, \gamma_t)$ and $u^P(s_t, d_t^F, \gamma_t)$ for agent and principal respectively are realized, where d_t^F is the final disposition of the claim.

In a way defined momentarily, the players' payoffs depend on the final resolution of the claim and whether the agent was reversed or affirmed. If the principal affirms, the final disposition matches the agent's disposition. If the principal reverses, the final disposition is the opposite of the agent's decision. To embed reputational concerns into the model, assume the agent may be one of two types: conscientious or biased. A conscientious agent always chooses as the principal would choose if she had access to the local information. The biased agent's preferences differ from the preferences of the principal. He acts strategically to maximize his utility given the strategy choice of the principal. The prior probability that an agent is conscientious is μ_0.

2.2 Preferences and Stategies

The strategy space of the principal is straightforward. On appeal, the principal observes x and the agent's resolution. Let $\gamma(x, d) \in [0, 1]$ be the probability that the principal affirms a decision d when the global fact is x.

The strategy space of the biased agent is potentially quite complex. The agent's strategy is a conclusion that the claim is valid or not valid for every claim in the unit square. In what follows, we restrict attention to cutline strategies of the form: $d(x, y) = \widehat{y}(x)$ where

$$d(x,y) = \begin{cases} v \quad if \quad y > \widehat{y}(x) \\ nv \quad otherwise \end{cases}$$

This class of strategies is quite reasonable; as we shall see, with these strategies, the agent mimics a conscientious agent when the loss to him is not too costly to him.[5] The principal infers the likely location of the local fact from the global fact, the agent's disposition, and her prior about the agent's type.

[5] A broader class of strategies would identify, for each x, some set $V(x) = \{y | d(x, y) = 1\}$ on which the agent set $d(x, y) = 1$; on the complement of $V(x)$, $d(x, y) = 0$. But the set V could be any subset of the unit interval; so one strategy would be to find validity when x is rational and no validity when x is irrational or to find validity when x is in the Cantor set but no validity when x is not in the Cantor set. These strategies, however, seem totally irrational given the structure of preferences of both agent and principal.

The principal and the agent have consequential preferences over the final resolution of the claim. Obviously, both principal and agent prefer a "correct" resolution to an "incorrect" resolution of the claim. Yet what is "correct" and what is "incorrect"? We suppose that each player has an ideal partition of the claim space into valid and invalid claims. A player views the resolution of claim s as "correct" if it conforms to the resolution dictated by her partition.[6]

The principal's ideal partition is defined by the line

$$x + y = 1$$

That is to say, for any claim, the principal prefers a valid resolution if and only if $x+y > 1$ and it prefers not valid otherwise. If the final resolution is correct from the principal's point of view, it receives 0. If the final resolution is incorrect, however, she loses more the "easier" the claim was to resolve – i.e. the further away the claim s is from the boudary of its set in the partition.[7] The preferences of the principal are thus represented by the following utility function:

$$u^P(x, y; d, \gamma) = \begin{cases} -|y - (1 - x)| & \text{if final and preferred resolution don't match} \\ 0 \quad \text{otherwise} \end{cases}$$

By contrast, the biased agent prefers a valid resolution if $y > \frac{1}{2}$. His preferences over claim resolution are:

$$\widehat{u}^A(x, y) = \begin{cases} -\left|y - \frac{1}{2}\right| & \text{if final and preferred resolution don't match} \\ 0 \quad \text{otherwise} \end{cases}$$

This specification of preferences makes the principal/agent conflict stark (and the algebra easier). The biased agent places relatively more weight on its local knowledge of the claim than the principal prefers. Suppose, as an example, the principal is Congress. The agent is an adminstrative agency. We might think of Congress as valuing expertise and politics in making a decision, whereas the agency values expertise alone. In short, the problem is that agents place too high a value on their own private information.

[6] Preferences are consequential in the sense that each court's utility is determined by the disposition of the case by the judicial system not by the individual court's own decision.

[7] More precisely, the principal has a linear loss utility function over resolutions a thorough discussion of possible preferences in this setting , see Cameron and Kornhauser (2017) which discusses the issue in the context of judicial decisionmaking..

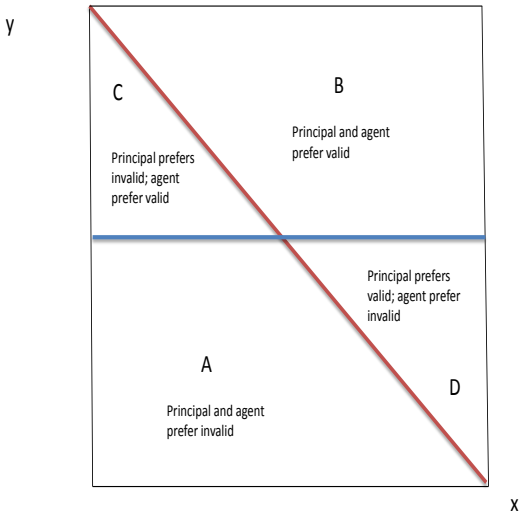

Figure 1: Preference Conflict

As figure 1 shows, the preferences of the principal and the biased agent are partially aligned. They agree about the resolution for claims in the areas A and B. They disagree about the resolution for claims in the areas C and D. Further, the extent of the preference conflict depends on the location of the global fact.

The agent's decision and the principal's review lead to a final disposition of the form:

$$d^F(d,\gamma) = \begin{cases} d & \text{if affirm } (\gamma = 0) \\ d^C & \text{if reverse } (\gamma = 1) \end{cases}$$

where, with some abuse of notation, d^C represents the decision not chosen by the agent.

Upon reversal, the agent suffers a loss of k. So the agent's full utility is:

$$u^A(s, d^F, \gamma) = \begin{cases} \widehat{u}^A(s,d) \text{ when } d^F = d \\ \widehat{u}^A(s,d^C) - k \text{ when } d^F = d^C \end{cases}$$

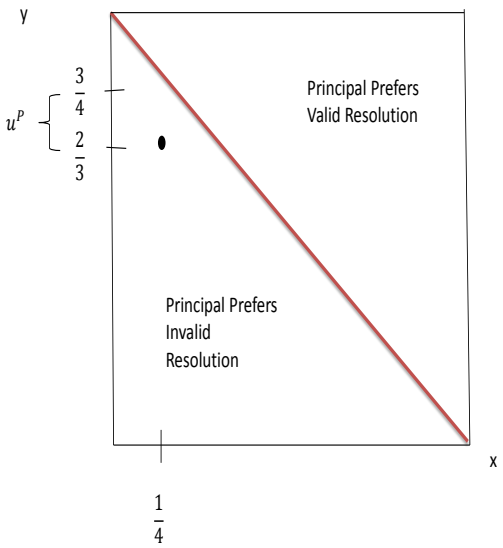

Figure 2: Principal's Disutility From Final Dispositions With Which He Disagrees

As an example, suppose that the claim is $s = \left(\frac{1}{4}, \frac{2}{3}\right)$. The agent determines that the claim is valid and the principal affirms. In that case, the final disposition is valid. The final disposition is correct from the agent's perspective. The principal also affirmed. As a result, the agent's utility is 0. On the other hand, the final disposition is incorrect from the principal's perspective. The principal's loss is

$$u^P = -(\frac{3}{4} - \frac{2}{3})$$

Figure 2 shows this disutility. The black dot is claim s. It falls below the 45 degree line; the principal prefers invalidity. The principal's loss from a valid final disposition is u^P – the vertical distance between his ideal partition at $x = 1/4$ and the claim.

Suppose instead that the agent determines this claim is valid and the principal reverses. In that case, the final disposition is "invalid." This disposition is correct from the principal's perspective and she suffers no loss. This disposition

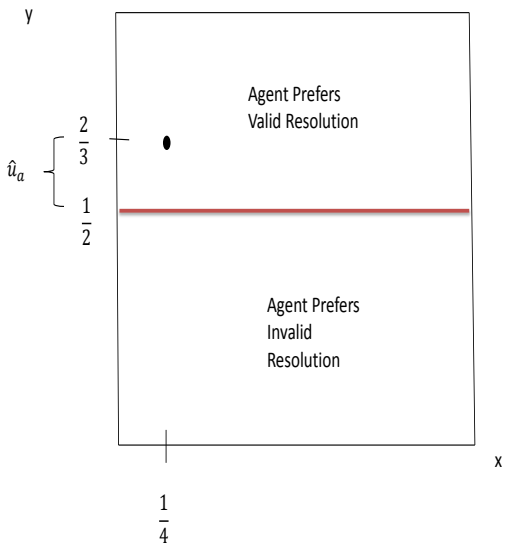

Figure 3: Agent's Disutility From Final Dispositions With Which He Disagrees

is incorrect from the agent's perspective, whose utility is

$$u^A = -(\frac{2}{3} - \frac{1}{2}) - k$$

Figure 3 shows the disutility associated with the incorrect disposition. Because of reversal, the agent suffers an additional cost k.

Our framework is simple. Though the agent may be of two types, the conscientious agent acts mechanically; she always resolves the claim as the principal would, were she acting alone with full knowledge. We thus need only consider the behavior of the biased agent. The equilibrium definition for the one-period benchmark follows:

Definition: Bayesian Nash Equilibrium in the One Period Benchmark:

The profile $\{y^*(x), \gamma_v^*(x, y^*(x)), \gamma_{nv}^*(x, y^*(x)), \mu^*\}$ *(where* $d^*(x) = 1$ when $y > y^*(x)$*) is a perfect Bayesian equilibrim if and only if (1) the biased agent's strategy maximizes its expected utility, given the strategy and beliefs of the principal; (2) the principal"s review strategy maximizes its expected utility following a valid or not valid resolution given its beliefs and the strategy of the biased agent; and (3) whenever possible, the beliefs of the principal derive from the equilibrium strategy of the biased agent according to Bayes rule.*

3 Equilibrium of the One Period Model

3.1 Easily Affirmed Claims

The model showcases both easy cases for affirmance and hard cases for affirmance. Consider claims in regions A and B in figure 1.[8] Here, as noted, the incentives of the principal and agent align. Further, the agent can cheaply signal the claim lies in either region.

Take claims with a global fact below 1/2. For these claims, the agent prefers to find more claims valid than the principal does. As a result, the agent's equilibrium strategy never involves setting the cutline above $1 - x$. An invalid claim thus arises when: (1) the claim lies in region A or (2) the claim lies below the principal's cutline. In either case, the principal wants to affirm the resolution.

The logic applies to claims with global facts above 1/2. For these claims, the agent prefers to find more claims invalid than the principal does. If the agent validates the claim either (1) the claim lies in region B or (2) the claim lies above the principal's cutline. Either way, the principal wants to affirm.

The simultaneous existence of easy and hard claims offer some intuitive results. The principal affirms any resolution that accords with what she expects

[8] Formally, region A is S^1=$\{(x, y)|x, y \leq .5\}$, region B is $S^2 = \{(x, y)|x, y \geq .5\}$, region C is $S^3 = \{(x, y)|x > .5$ and $y < .5\}$ and region D is $S^4 = \{(x, y)|x < .5$ and $y > .5\}$. Principal and agent agree on the appropriate resolution of claims that lie in $S^1 \cup S^2$.

given the location of x. If, say, the text of the contract suggests no liability and the trial court finds no liability, the appellate court affirms. On the other hand, if the text of the contract suggests liability and the trial court finds liability, the appellate court affirms. The principal only potentially reverses hard claims, claims where she isn't sure whether the disposition reflected biased behavior or extraordinary local facts.

3.2 Harder Claims to Affirm

Consider claims in regions C and D of figure 1.[9] For those claims, the biased agent may prefer a different resolution than the principal. To mitigate the agency cost, the principal might have to reverse the agent's resolution. Reversal, however, is potentially costly for the principal, as with positive probability, the "minority" (or "unexpected") resolution in the region may be the correct one.

Should the principal reverse an unexpected resolution? In these circumstances, the agent's resolution of the claim states that the persuasiveness of the unobserved fact y more than offsets the global fact x. There are two possibilities: (1) y is, in fact, so large that it more than offsets the value of x or (2) the agent is biased and y does not fully offset x. Here, the preference conflict between the principal and biased agent plays a role. Consider the situation of $x < 1/2$ and $x > 1/2$ in turn.

3.2.1 Unexpected Validity Claims

For unexpected valid claims, the principal's payoff from affirming is

$$-\int_{\widehat{y}(x)}^{1-x} (1-x-y) f(y|validity) dy$$

The principal's payoff from reversing is

$$-\int_{1-x}^{1} (1-x-y) f(y|validity) dy$$

Consider figure 4. The dotted line represents a possible equilibrium cutline for

[9] Formally, those areas are defined as $S^3 \cup S^4$.

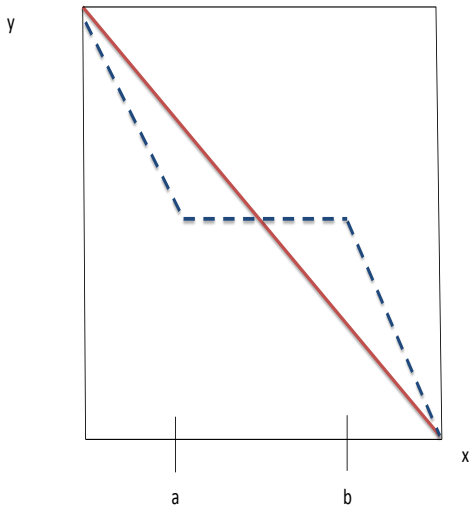

Figure 4: One Period Benchmark–Biased Agent's Equilibrium Strategy

the biased agent, $\widehat{y}(x)$. Under this strategy, the biased agent finds validity if the claim lies above the dotted line and invalid otherwise. Suppose the agent finds the claim valid and the global fact is less than 1/2. From this disposition, the principal knows that the claim must lie above the 45 degree line if the agent is concientious and above the agent's equilibrium cutline if the agent is biased. Suppose the principal affirms a valid resolution. The principal suffers a loss if the value of the local fact lies between $[y(x), 1-x]$ – the area between the 45 degree line and the equilibrium cutline for the biased agent. The extent of the loss depends on the distance between 45 degree line and the realization of y. Suppose instead the principal reverses. The final disposition is then invalid. The principal suffers a loss if the value of the local fact lies between $[1-x, 1]$ – the area above the 45 degree line. Again, the extent of the loss depends on the distance between the principal's preference line and the realization of y.

In this formulation, one can easily see the relationship between the equilibrium strategy of the biased agent and the principal's reversal decision. Suppose that the biased agent "pooled" He set his equilibrium strategy for every global fact equal to the principal's ideal partition. In that case, $f\ (y|valid) = 0$ for all values of $[1/2, 1-x]$. The principal would affirm all resolutions

More interesting, suppose that the biased agent sets $y^* = 1/2$. In figure 4 this is the horizontal part of the agent's equilibrium cutline. Will the principal nonetheless affirm the valid resolution – a resolution that goes against what the principal believes the correct answer should be, at least for some range of global facts? Given that equilibrium strategy, Bayes rule implies:

$$f(y|valid) = \begin{cases} \frac{1}{\mu_0 x + \frac{1}{2}(1-\mu_0)} & for\ y \in [1-x, 1] \\ \frac{1-\mu_0}{\mu_0 x + \frac{1}{2}(1-\mu_0)} & for\ \ y \in [\frac{1}{2}, 1-x] \end{cases}$$

The principal nonetheless affirms the unexpected resolution if

$$\frac{(1-\mu_0)\int_{1/2}^{1-x}(1-x-y)dy}{pr(valid)} - \frac{\int_{1-x}^{1}(1-x-y)dy}{pr(valid)} \leq 0 \qquad (1)$$

The positive root of equation (1) defines the lower bound of deference in the problem. The principal affirms all claims of validity when

$$x \geq a = \frac{z}{2}$$

where

$$z = \frac{\sqrt{1-\mu}}{1+\sqrt{1-\mu}}$$

For claims with global facts above a, the principal affirms whenever the agent finds validity.[10] Anticipating deference, for these global facts, the biased agent finds the claim valid whenever it prefers to do so. If the agent is certainly conscientious ($\mu = 1$), the principal delegates all decisionmaking ($a = 0$). When the agent is for certainly biased ($\mu = 0$), the principal delegates decisions where the global facts are located between $[1/4, 1/2]$. The range of discretion responds in a natural way to the belief about the agent's loyalty.

Next examine what happens when x lies below a. In that case, if the biased agent acted truthfully, the principal would reverse. On the other hand, if the agent "pooled," the principal would affirm. The equilibrium involves mixing. The principal reverses with positive probability ($\gamma^*(x, valid) \in (0, 1)$). Anticipating reversal, the biased agent partially pools, adopting a strategy $y^*(x) \in (1/2, 1-x)$ We derive the exact values in the proof of the first proposition below. Importantly, the probability of reversal increases as x moves towards zero. The reason resonates: Because the biased agent doesn't want to provoke reversal, it finds fewer and fewer claims of liability.

3.3 Unexpected Invalidity Claims

The principal's optimal review of unexpected invalidity claims mirror the discussion above. Suppose that the biased agent uses its preferred legal cutline. The principal nonetheless affirms the unexpected resolution of "not valid" if

$$\frac{(1-\mu_0)\int_{1-x}^{\frac{1}{2}}(y-(1-x))dy}{pr(not\ valid)} - \frac{\int_{0}^{1-x}(1-x-y)dy}{pr(not\ valid)} \leq 0 \tag{2}$$

Solving (2) yields the upper bound b. The principal affirms all claims of invalidity where $x < b = 1-a$. If the agent is certainly conscientious, the upper bound is 1. If the agent is certainly biased, then $b = \frac{3}{4}$. In the region of partial delegation ($x \in [b, 1]$), the principal mixes between reversing and affirming when it observes an unexpected disposition. The biased agent's strategy reflects partial compliance. Fearing reversal, the biased agent decides some (but not all) claims

[10] That is: for $x \in (a, .5)$, $\gamma_t^*(x_t, d_t^T(x, y), \mu_t) = 1$. i.e, the principal affirms with probability 1. Recall that, when $x < .5$, the principal affirms all claims resolved nv.

in a way he disfavors but the principal prefers. Figure 4 provides an example of the bounds a and b.

To sum up, the one period equilibrium consists of two regions with blurred bounds. For claims with relatively inconclusive global facts, the principal delegates completely. For claims where the global facts lie at the ends of the unit interval, there is partial delegation. Partial delegation means that, on the one hand, the principal affirms "expected" resolutions – resolutions that align with her expectations give realized the global fact. On the other hand, the principal reverses with positive probability "unexpected" resolutions – claims where the resolution and global fact are misaligned. Further, the degree of partial delegation decreases as the value of the the information that lies solely in the possession of the agent decreases. At the corners, there is no deference.

Having considered all the possible values of x, the first proposition specifies the equilibrium:

Proposition 1 *A one period equilibrium consists of three partitions:*

(1) ***Complete Deference****. If $x \in [a, b]$, the principal affirms all resolutions ($\gamma_{nv}^* = \gamma_v^* = 0$). The agent decides according to his preferred cutline ($y_t^*(x) = .5$). The principal's beliefs are $\mu(liable) = \frac{\frac{1}{2}(1-\mu_0)}{\frac{1}{2}(1-\mu_0)+\mu_0 x}$ and $\mu(not\ valid) = \frac{\frac{1}{2}(1-\mu_0)}{\frac{1}{2}(1-\mu_0)+\mu_0(1-x)}$.*

(2) ***Partial Deference to Valid Dispositions.*** *If $x \in [0, a]$, the principal affirms all invalid dispositions ($\gamma_{nv}^* = 0$) and reverses valid dispositions with probability $\gamma_v^* = \frac{(\frac{1}{2}-\frac{x}{z})}{(\frac{1}{2}-\frac{x}{z}+k)}$. The agent decides according to the cutline $y^*(x) = 1 - \frac{x}{z}$. The principal's beliefs are $\mu(valid) = \frac{(1-y^*(x))(1-\mu_0)}{(1-y^*(x))(1-\mu_0)+\mu_0 x}$ and $\mu(not\ valid) = \frac{y^*(x)(1-\mu_0)}{y^*(x)(1-\mu_0)+\mu_0(1-x)}$.*

(3) ***Partial Deference to Invalid Dispositions****. If $x \in [b, 1]$, the principal affirms all valid dispositions ($\gamma_v^* = 0$) and reverses invalid dispositions with probability $\gamma_{nv}^* = \frac{\left[\frac{1}{2}-\frac{(1-x)}{z}\right]}{\left[\frac{1}{2}-\frac{(1-x)}{z}+k\right]}$. The agent decides according to the cutline $y^*(x) = \frac{1-x}{z}$. The principal's beliefs are $\mu(valid) = \frac{(1-y^*(x))(1-\mu_0)}{(1-y^*(x))(1-\mu_0)+\mu_0 x}$ and $\mu(not\ valid) = \frac{y^*(x)(1-\mu_0)}{y^*(x)(1-\mu_0)+\mu_0(1-x)}$.*

Proof

As derived in the text, the cutlines for deference are

$$a = \frac{z}{2}$$

and

$$b = 1 - a = 1 - \frac{z}{2}$$

Consider cases with global facts such that $x < a$. At the agent's cutline strategy $(\widehat{y})$, the agent must be indifferent between validating the claim and not, given the probability of reversal of a validity claim, $\widehat{\gamma}_v$. The agent's indifference expression is

$$-(\widehat{y} - \frac{1}{2}) + \widehat{\gamma}_v(\widehat{y} - \frac{1}{2} + k) = 0 \tag{3}$$

At the same time, the principal must be indifferent between reversing and affirming, given beliefs consistent with the strategy $\widehat{y}$. As a result, it must be that

$$x - (1 - \widehat{y})z = 0 \tag{4}$$

The equilibrium is defined as the joint solution to (3) and (4) along with beliefs $\mu = \frac{\mu_1 x}{\mu_1 x + (1-\mu_1)(1-y^*)}$. Solving equation (4) for $\widehat{y}$ yields

$$\widehat{y} = 1 - \frac{x}{z} \tag{5}$$

Replacing $\widehat{y}$ in (3) gives

$$-(\frac{1}{2} - \frac{x}{z}) + \widehat{\gamma}_v(\frac{1}{2} - \frac{x}{z} + k) = 0$$

or

$$\widehat{\gamma}_v = \frac{(\frac{1}{2} - \frac{x}{z})}{(\frac{1}{2} - \frac{x}{z} + k)} \tag{6}$$

Since $a = \frac{1}{2}z$, if $x < a$, equation (6) gives a positive value. The equilibrium values γ_v^*, y^* are given by the solutions to (5) and (6).

Consider cases with global facts $x > b$. At the agent's cutline strategy $\widehat{y}$, the agent must be indifferent between invalidating the claim and not, given the probability of reversal of invalidity $\widehat{\gamma}_{nv}$. The agent's indifference equation is

$$-(\frac{1}{2} - \widehat{y}) + \widehat{\gamma}_{nv}(\frac{1}{2} - \widehat{y} + k) = 0 \tag{7}$$

At the same time, the principal must be indifferent between affirming and reversing, which requires that

$$\widehat{y} = \frac{1 - x}{z} \tag{8}$$

The equilibrium is the joint solution to (7) and (8) along with beliefs $\mu =$

$\frac{\mu_1(1-x)}{\mu_1(1-x)+(1-\mu_1)y^*}$.

Replacing $\widehat{y}$ in equation (7) with the expression in equation (8) gives

$$-\left[\frac{1}{2}-\frac{(1-x)}{z}\right]+\widehat{\gamma}_{nv}\left[\frac{1}{2}-\frac{(1-x)}{z}+k\right]=0$$

Solving for $\widehat{\gamma}_{nv}$ yields

$$\widehat{\gamma}_{nv}=\frac{\left[\frac{1}{2}-\frac{(1-x)}{z}\right]}{\left[\frac{1}{2}-\frac{(1-x)}{z}+k\right]} \tag{9}$$

Recall that $(1-b)=\frac{z}{2}$. As a result for $x>b$, we have $(1-b)<\frac{z}{2}$ and $\widehat{\gamma}_{nv}$ is positive. The equilibrium values γ^*_{nv}, y^* are the solutions to (8) and (9) ■

This proposition identifies the principal's trade-off. For $x \in [a,b]$, the principal benefits from the agent's local knowledge but the agent exploits his superior knowledge to resolve claims according to his own preferences. The location of a and b reflects both the principal's beliefs about the likelihood that the agent is biased and the probability that the claim will be unfavorably decided by the biased agent.

The model sheds light on a number of aspects of claim resolution. First, it demonstrates the importance of red flags. Resolutions that go against the principal's instincts raise red flags. The scope of the flag is determined by an estimate as to whether the local information overwhelmed the global fact. What are the chances that the star witness testimony was so convincing that his testimony alone overwhelmed the clear text of contract, making the right choice liability? The agent understands which resolutions raise red flags and responds optimally.

Consider the resolution of habeas petitions by a federal district court. The appellate court expects most, if not all, these petitions will be denied. Thus the denial of a petition is never reversed. On the other hand, suppose the trial court grants the petition. That decision raises a red flag for the appellate court. Nonetheless, the appellate court might still affirm if, given what it can easily observe, the case looks close.

Second, principals often uphold reasonable decisions even if they think the decision is wrong. Why? The model formalizes standards of review like "manifestly errorneous," "clear error," and "reasonable." The deference interval $[a,b]$ determines the range of reasonable. For claims in these interval, the principal has a conjecture about the right answer. The principal nonethless defers to all resolutions. In other words, she concludes that the agent has acted reasonably,

no matter the resolution. A manifest error, by contrast, occurs when the global fact conflicts sufficiently with the resolution.

Third, unlike many models of delegation (get cites), the agent's faces two bounds on its decision, not one. The agent doesn't just face, say, an inflation cap or target. Instead, the agent's decisions below and above certain thresholds are potentially reversed. The reasons are twofold: First, a claim lies in two dimensions, only one of which is private information. Second, the preference conflict involves the relative weight placed on local and global information. In the standard delegation and cheap talk models, the agent prefers "more action" in every state. The agent doesn't dispute the relative value of what he alone knows.

We close this section with a legal application. Consider a tort claim. The law states that the defendant should be liable (i.e. the plaintiff's claim is valid) if he acted negligently. In a bench trial, the trial court/agent first finds facts. It then asks what law governs the defendant's conduct (i.e., negligence). Finally, it applies the law to the facts, asking whether the facts it found constitute negligence. The black letter law for the principal's review of this decision is simple. The principal reviews finding of fact for clear error. In this example, the finding of fact is the agent's report about x and y. The principal reviews findings of law de novo. The finding of law is that a negligence standard governs tort actions. The principal's review of the application step is more muddy. The principal reviews mixed questions of law and fact (was the defendant negligent? Was the police officer's search reasonable?) on a sliding scale. If the inquiry is primarily factual, it applies a standard of "clear error". If the inquiry is primarily legal, the standard of review is *de novo*. But what makes something "more factual" or "more legal."? The Supreme Court has repeatedly stated that the difference between law and fact cannot be gleaned from the caselaw (cites).

The model provides a way to think about what is going on. The appellate court doesn't know y. The trial court might report a large finding of fact to justify its imposition of liability. Deference to the trial court's resolution (the application of law to fact) turns on how important the principal thinks that finding is. If critical, the principal defers, even if the resolution seems suspect based on the global fact. If not critical, the principal reverses with some positive probability. The end result is blurry discretionary bounds, a finding in accord with the Supreme Court's statement about the law/fact distinction. Inside the deference bounds, delegation is complete. Outside the bounds, delegation is partial, turning on the "expectedness" of the resolution.

4 Two Period Model

This section analyzes the two-period model. In this model, the principal can alter the bounds of discretion over time. We study the relationship between the claim presented in period one and the agent's reputation in period 2. What claims make or break the agent's reputation; in other words, what claims are the most effective tests of loyalty? We also will compare the bounds in the first period of the dynamic game to the equilibrium bounds in the one-period model. Several questions arise. First, does the addition of the second period induce the principal to allocate greater discretion to the agent in the first period? What does the anticipation of career concerns by the agent do to its decision in period 1? Third, does a successful agent (an agent whose first period disposition matched the principal's desires) always earn more discretion? The principal cannot fire the agent. The method of control is a (credible) threat to reverse with higher frequency, in effect allocating fewer and fewer matters to the agent's sole discretion.

The principal learns at the close of period 1 whether the agent's resolution resulted in a loss to her. The principal uses this information to update her beliefs about the agent. As the conscientious agent always acts according to the principal's preferences, each time the principal observes a loss or failure from the period one resolution, it learns that the agent is biased. On the other hand, success is more ambiguous. If the principal observes success, it could be that the agent is conscientious or it could be that the biased agent heard a claim that he preferred to decide as the principal would.

In the two period model, the principal updates with each new piece of evidence. She updates following the period 1 disposition. At that moment, the principal must decide whether to reverse or affirm. The principal updates again following the realization of her period 1 payoff. At that moment, she knows whether she suffered a loss or not. The principal, for example, might reverse the agent's disposition and later learn that reversal was the wrong move. In that circumstance, the principal regrets the final disposition. In acknowledging that she was wrong and the agent was right, the principal increases her assessment that the agent is a conscietious type.

Denote as μ_1 the principal's belief that the agent is good following the period 1 disposition. Denote as $\mu_{1.5}$, the principal's belief following the realization of her period one payoff but before the period 2 disposition. Finally, denote as μ_2, the principal's belief following the period 2 disposition. The equilibrium

definition for the two period model follows:

Definition: Bayesian Nash Equilibrium in the Dynamic Game

A profile $(y_1^*, y_2^*, \gamma_1^*, \gamma_2^*, \mu_1^*, \mu_{1.5}^*, \mu_2^*)$ *forms a perfect Bayesian equilibrim if (1) the biased agent"s strategy maximizes his current and future expected utility at each point in time, given the strategy and beliefs of the principal; (2) the principal's strategy at each point in time maximizes her current and future utility given her beliefs and the strategy of the biased agent; and (3) whenever possible, the beliefs of the principal are derived from the equilibrium strategy of the biased agent and the unmodeled behavior of the conscientious agent according to Bayes rule.*

In period 2, the biased agent's expected loss turns on the amount of discretion, which in turn depends on the principal's beliefs. The equilibrium is, adjusted for the updated beliefs, as described in the one-period model. For beliefs, μ, the biased agent's second period expected payoff is

$$W(\mu) = -2\left\{ \int_0^{a_2(\mu)} \int_{\frac{1}{2}}^{y_2^*(x)} (y_2 - \frac{1}{2}) dy_2 dx_2 + \int_0^{a_2(\mu)} \int_{y_2^*(x)}^{1} \gamma_v(x)(y_2 - \frac{1}{2} + k) dy_2 dx_2 \right\}$$

Figure 5 provides the intuition. For claims arising in the region of complete discretion, the biased agent suffers no loss. For claims below $a_2(\mu)$, the biased agent suffers a loss in two circumstances: First, a claim might arise that the biased agent decides not valid rather than valid when it prefers not to (claims where $x > .5$ and $y \in [1/2, y_2^*(x)]$. It does so to avoid provoking reversal. That is the value of the first integral (and the claims in blue area in figure 5). Second, the biased agent might find the defendant valid and suffer reversal with some positive probability. That is the value in the second integral (and the claims in the green area in figure 5, discounted by the value $\gamma_v(x)$). The expression is symmetric on the top end on the interval, leading to a total loss of the lower interval multiplied by two.

Take a step back and consider the biased agent's behavior in period 1. Suppose that the principal fully allocates discretion. In the one period model, this leads to biased behavior on the part of the agent. He decides according to his preferred rule. In the two period model, career concerns come into play. If the agent decides according to his preferred rule, the principal will suffer a loss for some claims, revealing the agent's type. The principal will update her belief to $\mu = 0$ and contract the range of discretion in period 2. Reputational concerns

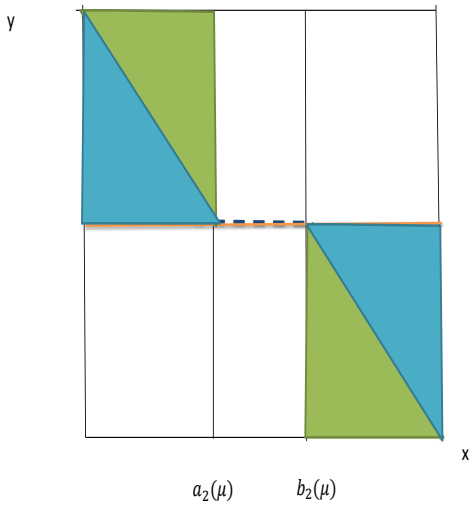

Figure 5: Agent's Period 2 Expected Utility

might cause the agent to mimic the behavior of the conscientious agent.

Suppose, as an example, that the biased agent observes a claim where $x < \frac{1}{2}$ and $y \in [\frac{1}{2}, 1-x]$. In such a claim, the agent prefers to hold the claim valid but the principal prefers invalidity. What happens if the biased agent finds the claim valid? Before the next period, the principal will observe a loss and learn that the agent was biased. It will thus allocate less discretion in the second period. Rather than risk this chain of events, the biased agent might simply follow the principal's preferred disposition. The payoff from finding no validity (and hiding oneself) is

$$-(y - \frac{1}{2}) + W(\mu_0)$$

For the same claim, the payoff from finding validity (and revealing oneself) is

$$W(0)$$

The agent prefers to pool in period 1 if

$$-(y - \frac{1}{2}) + W(\mu_0) \geq W(0) \tag{10}$$

By inspection of equation (10), one can see that for claims with global facts relatively close to 1/2, the biased agent pools. Take global fact: $1/2-\varepsilon$. Suppose that the agent pools and sets its cutline at $1 - x_1$. Finally, suppose that a case of "conflict" arises; that is to say, $y \in [\frac{1}{2}, 1/2 + \varepsilon.]$. Upon following her pooling strategy, the agent suffers a first period loss of on this case of ε, but she preserves her reputation ($W(\mu_0)$ instead of $W(0)$).[11] Suppose instead the agent followed her strategy from the one period model, acting myopically. She sets the cutline at $\frac{1}{2}$. The same case of conflict arises. Under the myopic strategy, the agent suffers no first period loss, but suffers a dramatic contraction in discretion in period 2. Given these two choices, the agent prefers the pooling cutline.

For global facts at the corners, the reputation effects are too meager to induce sufficient compliance by the agent. For these claims, the principal needs an additional lever of control. So, she reverses unexpected dispositions with positive probability. For claims between the perfect pooling region and the reversal region, the agent balances the reputational hit against the short term

[11] Note that for cases where $y < \frac{1}{2}$ and $y > 1 - x$, there is no conflict between the agent's short run and long run goals. For cases where $y < \frac{1}{2}$ she decides invalid, the principal suffers no loss and the agent preserves her reputation. For cases where $y > 1 - x$, the agent decides the claim as valid, the principal suffers no loss and the agent preserves her reputation. The only claims that present a long run short run tradeoff are those in the region of conflict.

loss from deciding a case against her own interest. Given the equilibrium cutline induced by reputational concerns, the principal affirms. Placing this altogether, the next proposition sets forth the two period equilibrium.

Proposition 2 *In the two period model, an equilibrium is characterized as follows:*

1. *In interval* $[0, a_1]$*, the principal affirms all not valid resolutions, it reverses valid resolutions with probability* γ_l^**, where* γ_l^* *solves the expression (A1) in the appendix and the agent"s strategy makes equation (A2) hold with equality. In the interval* $(a_1, \alpha_1]$*, the principal affirms all resolutions, the agent's strategy solves expression (A3). In the interval* $(\alpha_1, \beta_1]$*, the principal affirms all resolutions and the agent perfectly pools (sets* $y_1^* = 1 - x$*). In the interval* $(\beta_1, b_1]$*, the principal affirms all resolutions and the agent's strategy solves expression (A4). Finally, in the interval* $(b_1, 1]$*, the principal affirms all valid resolutions; she reverses invalid dispositions with probability* γ_{nl}^**, where* γ_{nl}^* *solves expression (A5).*

2. *Dispositional Beliefs (Period 1). The principal's beliefs following a valid resolution are* $\mu_1 = \frac{\mu_0 x_1}{\mu_0 x_1 + (1-\mu_0)(1-y_1^*)}$*. The beliefs following a not valid resolution are* $\mu_1 = \frac{\mu_0(1-x_1)}{\mu_0(1-x_1)+)(1-\mu_0)y_1^*}$*.*

3. *Final Beliefs (Period 1): The principal's beliefs after suffers a loss following the period 1 final disposition are* $\mu_2^F = 0$*. When* $x_1 < \frac{1}{2}$*, the principal's belief after sufffering no loss from a valid disposition in period 1 is* $\mu_2^S = \frac{\mu_0}{\mu_0 + (1-\mu_0)(x_1 + y_1^*)}$*. When* $x_1 > \frac{1}{2}$*, the principal's belief after suffering no loss from an* $\mu_2^S = \frac{\mu_0}{\mu_1 + (1-\mu_0)((1-x_1)+(1-y_1^*))}$*.*

4. *Equilibrium (Period 2). The equilibrium in period 2 is as described in proposition 1 where the principal's belief about the agent's type is* μ_2^S *or* μ_2^F*, depending on the outcome in period 1.*

Proposition 2 identifies how reputation and learning affect the structure of equilibrium. Figure 6 represents the agent's equilibrium cutline in period 1.Compare this equilibrium to a comparable game in which the principal plays the static game with twice, each time with a different agent. In this simple situation, the principal selects the same bounds for each agent and each agent exercises his discretion to decide each claim within the interval $[a, b]$ as he thinks best. Compare this outcome first to the behavior in period 2 of the dynamic

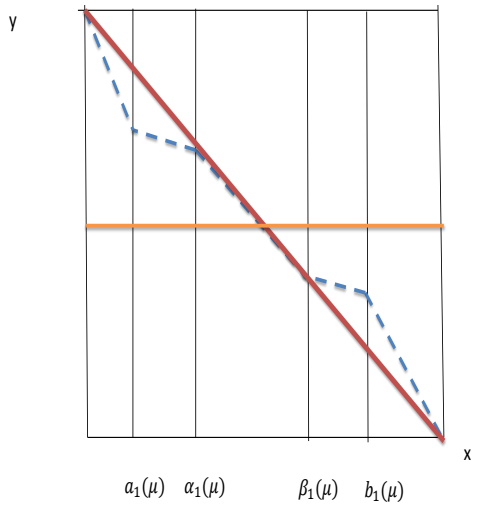

Figure 6: Agent's Period 1 Equilibrium Cutline Utility

game. Notice first that behavior in the second period depends on the realization of the claim $s_1 = (x_1, y_1)$ in period 1 because, after she reviews the agent's decision, the principal learns whether the agent decided the claim contrary to the principal's views.

After period 1, the principal either knows with probability 1 that the agent was biased; when this event occurs, we shall say that that the agent's bias has been revealed (or unmasked) The principal can learn that the agent was biased in one of two ways. First, the principal can affirm the agent's incorrect decision. This results in a loss to the principal. Thus, from an affirmance followed by a loss, the principal can infer that the agent is biased. Second, the principal can reverse an agent's incorrect decision; this reversal does not incur a loss. Nevertheless, from observing reversal and no loss, the principal can infer that the agent initial decision was incorrect. In each event, the principal learns that, with probability 1, the agent was biased. Accordingly, she sets the period 2 bounds at $[a_2, b_2] = [.25, .75] = B_0$. In the other two events – affirmance of a correct decision without a loss or reversal of a correct decision with a loss, the principal infers that the agent decided consistently with her preferred decision. Consequently, she updates her beliefs that the agent is conscientious. The principal in period 2 thus allows the agent more discretion than she receives in the static game; i.e. $[a_2, b_2] \supset [a, b] = B_{static}$.

Notice that bounds can either increase or decrease after both an affirmance and a reversal. The principal can learn from either action. Thus, when she incurs a loss after a reversal, she *rewards* the agent by increasing the delegation to the agent just as, upon incurring a loss after an affirmance, she *punishes* the agent by restricting the delegation to the agent in period 2.

We may summarize the principal's behavior in period 2 as follows. When the principal learns the agent's type at the end of period 1, she tailors the period 2 bounds of discretion to the revealed type. On the other hand, when the agent's type is not revealed in period 1, the period 2 bounds will be broader than they were in the static model and that implies that, in period 2, the principal's losses may be greater in the dynamic game. Despite these competing effects, the next proposition shows that the principal is always better off when the agent is long-lived.

Proposition 3 *The principal's expected utility is strictly higher when she faces one agent that lives two periods than two different agents, each one living one period.*

Return to the equilibrium identified in proposition 2. We now consider the behavior of the agent in period 1. Note that the threat of narrow bounds in period 2, induces the agent in period 1 to mimic a conscientious agent over interval $[\alpha_1, \beta_1]$. Even though the agent has complete discretion to decide as he wishes, he resolves claims as the completely informed principal would. In period one, then, the principal is unambiguously better off over this agent than she was in the static, one-period model as, over the interval $[\alpha_1, \beta_1]$ she suffers no expected losses in the dynamic model though she does in the static model.

Moreover, because period 1 might reveal the agent's bias which would provide gains to the principal in period 2, the principal has an incentive to increase the bounds in which she affirms all of the agent's resolutions. In parts of this region – i.e., in $[a_1, \alpha_1]$ and in $[\beta_1, b_1]$, the agent decides according to his preferences. The next proposition shows that $[a_1, b_1] \supset [a, b]$.

Proposition 4 *The interval of global facts where the principal affirms all of the agent's decisions is larger in the first period of the two period model than in the one period benchmark.*

Proof. The lower bound in the benchmark is

$$a = \frac{1}{2}z$$

The lower bound of in the first period of the two period model is

$$a_1 = z(1 - y^*)$$

From the proof of proposition 2, we know that $y^c = y^*$ at the point a_1. Further, proposition 2 shows that $y^c > \frac{1}{2}$ when $x_1 < \frac{1}{2}$. And so,

$$a = \frac{1}{2}z > z(1 - y^c) = z(1 - y^*) = a_1$$

The upper bound in the benchmark is

$$b = 1 - \frac{1}{2}z$$

The upper bound in the first period of the two period model is

$$b_1 = 1 - z(1 - y^*)$$

From the proof of proposition 2, at b_1 we know that $y^c = y^*$. Thus, at

$$b_1 = 1 - z(1 - y^*) > 1 - \frac{1}{2}z$$

■

Recall our earlier discussion of the revelation and concealment of the biased agent. In the static model, in the region $[a, b]$ in which the resolution is delegated to the agent, the revelation of bias occurs only when the principal realizes a loss at the end of the game. In the dynamic model, the resolution is delegated to the agent in the region $[a_1, b_1]$ but bias can never be revealed in the interval $[\alpha_1, \beta_1]$ in which the biased agent mimics the conscientious agent. In the region $[a_1, \alpha_1] \cup [\beta_1, b_1]$, the revelation of a biased agent can occur only after the principal realizes a loss at the end of period 1. In both models, outside of the region of delegation, a biased agent can be revealed when the principal either affirms or reverses the agent's resolution of the claim. The reputational effect of observing a succcessful disposition (one where the principal suffers no loss) depends in non-obvious ways on the location of the global fact. It isn't true that the claims where bias is more prevalent generate the largest reputational boost. ■

Proposition 5 *The relationship between success and failure and agent's reputation in period 2 is as follows: (A) No matter the location of the global fact in period 1, when the agent's bias is revealed at the end of period 1, the second period discretion bounds are constant (*$[1/4, 3/4]$*); (B) When the agent's bias remains hidden after period 1, if the first period global fact lies in the pooling interval –* $[\alpha_1, \beta_1]$ *– success in period 1 does not change the bounds in period 2. (C) For global facts in the interval* $[0, \alpha_1]$*, the reward for success in period 1 increases and then decreases* x_1*. (D) For the global fact in the interval* $[\beta_1, 1]$*, the reward for success in period 1 increases then decreases in* x_1*.*

Intuition suggest that principals should reward success by granting greater discretion going forward. In this model, that doesn't happen. The reason runs as follows: For some claims, the principal expects success from every agent. When success is expected, its realization doesn't provide new information. It is like giving a test where every student (good or bad) receives an A.

The model also shows that it is much easier to lose a reputation than to gain one. The concientious agent follows the preferred strategy of the principal. Thus, each time that the pricnipal observes failure, she knows the agent is biased. Success is a mixed bag. It could be either type of agent. The biased

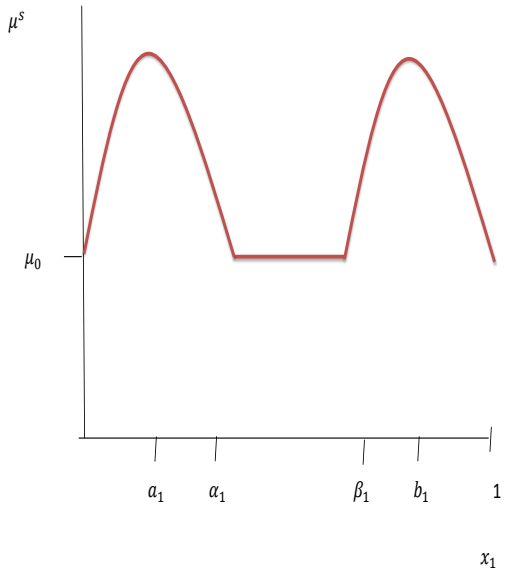

Figure 7: Reputational Benefits of Success

agent on occasion makes a voluntary decision to lose his reputation. She forfeits her credibility in exchange for the short term gain associated with deciding a claim in period one the way she prefers.

Third, one might think that success on "harder" tests of loyalty should correspond to greater future rewards. One natural way to think of a "hard" loyalty test is a claim with the most severe conflict between principal and agent preferences. Yet, in this model, the reputational boost from success is not the highest for claims where $x_1 = 0$ or $x_1 = 1$. In these cases, the agent is forced to pool out of a fear of reversal (and suffering the corresponding cost of k). Further, although these claims are potentially the most informative, the principal can't commit to affirm an unexpected disposition. She reverses with positive probability.

Figure 7 illustrates the relationship between the principal's second period beliefs and a successful resolution by the agent in period 1, given a claim with global fact x.

5 Concluding Remarks

Principals often delegate decisions to agents. In many, but not all, such contexts, private parties can commit to this delegation through contract. Generally, in the public sector, by contrast, the principal cannot commit to her delegation through an enforceable contract. In public bureacuracies, this inability to commit derives from the limitations on the employment contract. In the judiciary, though a hierarchy of courts exist, there are no mechanisms of control other than affirmance and reversal of the decisions of the lower court. An appellate court, thus, cannot commit to defer to the decisions of a lower court or an administrative agency. Our analysis shows that, nonetheless, when the agent has private information that is valuable to her, the principal will rationally delegate decisions to the agent. She will defer to the decision of the agent over a suitable region.

We then showed how, in a dynamic game, the principal may learn whether an agent is biased or conscientious. When learning occurs, the bounds in which decisions are delegated shift between periods. These bounds may either expand or contract. Moreover, the threat of contracting bounds (and the promise of expanding bounds), induces a biased agent to mimic a conscientious agent, thereby improving the payoff to the principal.

References

[1] Athey S., Atekson, A. & Kehoe, P. 2005. The Optimal Degree of Discretion in Monetary Policy. Econometrica, 73(5), 1431-1475.

[2] Amador, M. & Bagwell, K. 2013. The Theory of Optimal Delegation with an Application to Tariff Caps. Econometrica, 81(4), 1541-1599.

[3] Cameron, C. & Kornhauser, L. 2017. What Do Judges Want?, Chapter 3, mimeo, April 4, 2017.

[4] Crawford Vincient C. & Sobel, J. 1982. Strategic Information Transmission. Econometrica, 50(6), 1431-1451.

[5] Daughety, A. & Reinganum, J. 2000. Appealing Judgments. Rand Journal of Economics. 31(3). 502-525.

[6] Holmstrom, Bengt. 1979. Moral Hazard and Observability, Bell Journal of Economics, 10(1), 74-91.

[7] Holmström, Bengt. 1984. On the Theory of Delegation. In Bayesian Models in Economic Theory, edited by Marcel Boyer and Richard E. Kihlstrom, Amsterdam: North-Holland, 115-141.

[8] Huber K. & Shipan, C. 2006. Politics, Delegation, and Bureacracy. In B. Weingast and D. Wittman (eds) The Handbook of Political Economy (Oxford: Oxford University Press), pp. 256-272.

[9] Kornhauser, L. 1992. Modeling Collegial Courts. II. Legal Doctrine. Journal of Law, Economics, & Organization.8(3) 441–470.

[10] Lax, J. 2012. Political Constraints on Legal Doctrine: How Hierarchy Shapes the Law. Journal of Politics. 74(3). 765-781.

[11] Morris, S. 2001. Political Correctness. Journal of Political Economy, 109(2), 231-265.

6 Appendix: Proofs

Proof of Proposition 2

To prove proposition 2, the following two lemmas are useful. The first lemma identifies the highest and lowest utility the agent can receive in the second

period for any level of reputation. In this way, it establishes the "maximum" reputational punishment. The first lemma also establishes that the agent's second period payoff increases in his reputation. The second lemma shows that the numbers defining the partition of the global facts for different equilibria exist and are unique. Let μ_2 denote the principal's beliefs entering into period 2.

Lemma 6 *(a) For any value of $\mu_2 \in [0,1)$, the agent's second period payoff, $W(\mu_2)$, must be less than $\overline{W} = 0$; (b) For any value of $\mu_2 \in [0,1)$, the agent's second period payoff, $W(\mu_2)$ must be greater than $\underline{W} = -\frac{1}{24}$; (c) For any value of $\mu_2 \in [0,1)$, the agent's second period payoff increases in his reputation ($W'(\mu_2) > 0$).*

Proof:

(a) With a perfect reputation, proposition (1) teaches that $a_2 = 0$ and $b_2 = 1$. Thus, $W(1) = \overline{W} = 0$.

(b) Suppose that the principal set $a_2 = b_2 = \frac{1}{2}$. Further suppose the principal could commit to reverse all decisions outside the bounds. The payoff to such a plan must be worse (indeed the worst possible) for the agent. Under that strategy, we have

$$\underline{W} = -2\int_0^{\frac{1}{2}}\int_{\frac{1}{2}}^{1-x_2}(y_2 - \frac{1}{2})dy_2dx_2 = -2\int_0^{\frac{1}{2}}\frac{\left[\frac{1}{2} - x_2\right]^2 dx_2}{2} = -\frac{\left(\frac{1}{2}\right)^3}{3} = -\frac{1}{24}$$

(c) We want to show that the agent's second period payoff, $W(\mu_1)$,increases in μ_1. Recall that

$$W(\mu) = -2\left\{\int_0^{a_2(\mu)}\int_{\frac{1}{2}}^{y_2^*(x_2)}(y_2 - \frac{1}{2})dy_2dx_2 + \int_0^{a_2(\mu)}\int_{y_2^*(x)}^{1}\gamma_v(x)(y_2 - \frac{1}{2} + k)dy_2dx_2\right\}$$

Define the variable

$$z_1 = \frac{\sqrt{1-\mu_1}}{1+\sqrt{1-\mu_1}}$$

After this change of variable, write the agent's payoff $W(z_1(\mu_1))$. Notice that:

$$\frac{dz_1}{d\mu_1} = \frac{-\frac{1}{2}}{(1+\sqrt{1-\mu_1})^2} < 0$$

So the sign of $\frac{dW}{d\mu_1}$ is minus the sign of $\frac{dW(z_1((\mu_1))}{dz_1}$.

In the equilibrium of the second period, we have

$$y_2^*(x_2) = 1 - \frac{x_2}{z_1} \tag{11}$$

and

$$a_2(z) = \frac{z_1}{2}$$

Replace the value of $y_2^*(x_2)$ given above in $W(\mu_1)$:

$$W(z_1) = -2 \int_0^{a_2(z_1)} \int_{\frac{1}{2}}^{1-\frac{x_2}{z_1}} (y_2 - \frac{1}{2}) dy_2 dx_2 + \int_0^{a_2(z_1)} \int_{1-\frac{x_2}{z_1}}^{1} \gamma_v(x_2)(y_2 - \frac{1}{2} + k) dy_2 dx_2$$

The derivative consists of two terms (the bounds of discretion depend on z_1 and the equilibrium cutline in y space also depends on z_1).

$$\frac{dW}{dz_1} = -2 \left[\int_{\frac{1}{2}}^{1-\frac{a_2}{z_1}} (y_2 - \frac{1}{2}) dy_2 + \int_{1-\frac{a_2}{z_1}}^{1} \gamma_v(x)(y_2 - \frac{1}{2} + k) dy_2 \right] \frac{da_2}{dz_1}$$
$$- 2 \int_0^{a_2(z)} [(1 - \frac{x_2}{z_1} - \frac{1}{2}) - \gamma_v(x_2)(1 - \frac{x_2}{z_1} - \frac{1}{2} + k)] \frac{x_2}{z_1^2} dx_2$$

Note, however, that

$$\gamma_v(x) = \frac{y_2^*(x_2) - \frac{1}{2}}{1 - \frac{x_2}{z_1} - \frac{1}{2} + k}$$

Thus, the integrand of the second term equals 0 for all values of x_2. As a result,

$$\frac{dW}{dz_1} = - \left[\int_{\frac{1}{2}}^{1-\frac{a_2}{z_1}} (y_2 - \frac{1}{2}) dy_2 + \int_{1-\frac{a_2}{z_1}}^{1} \gamma_v(x)(y_2 - \frac{1}{2} + k) dy_2 \right] < 0$$

Replace $a_2 = \frac{z_1}{2}$ in the bounds of integration. The first integral goes away. And we have

$$\frac{dW}{dz_1} = - \int_{\frac{1}{2}}^{1} \gamma_v(x)(y_2 - \frac{1}{2} + k) dy_2$$

It is immediate that

$$\frac{dW}{d\mu_1} = \frac{dW}{dz_1}\frac{dz_1}{d\mu_1} > 0$$

■

Lemma 7 *(a) Define two functions of x_1 as*

$$H(x_1) = -(\frac{1}{2} - x_1) + W(\mu_0) - W(0)$$

and

$$G(x_1) = -\left(y^*(x_1) - \frac{1}{2}\right) + W(\mu^S(y^*(x_1), x_1)) - W(0) \tag{12}$$

where

$$\mu^S(y^*(x_1), x_1) = \frac{\mu_0}{\mu_0 + (1-\mu_0)(x_1 + y^*(x_1))}$$

and

$$y^*(x_1) = 1 - \frac{x_1}{z_1}$$

The claim is that (i) there exists numbers $\alpha_1 \in [0, \frac{1}{2}]$ and $a_1 \in [0, \alpha_1]$ that solve the expressions

$$H(\alpha_1) = 0 \tag{13}$$

and

$$G(a_1) = 0$$

(ii) The derivatives of H and G are positive over the entire interval. As a result, $H(x_1) < 0$ if $x_1 < \alpha_1$ and $H(x_1) > 0$ if $x_1 > a_1$. Likewise, $G(x_1) < 0$ if $x_1 < \underline{a}_1$ and $G(x_1) > 0$ if $x_1 > \underline{a}_1$

Proof :

Apply the intermediate value theorem to the function $H(x_1)$. Notice that $H(1/2) > 0$. From Lemma 1, we know that

$$\overline{W} - \underline{W} > W(\mu_0) - W(0)$$

As a result:

$$H(0) = -\frac{1}{2} + W(\mu_0) - W(0) < -\frac{1}{2} + \overline{W} - \underline{W} = -\frac{1}{2} + \frac{1}{24} < 0$$

Finally, differentiation shows that $H' > 0$. Thus, equation (13) defines a number lies between $[0, 1/2]$.

(b) Replacing y^* in expression (**??**) with its value in expression (13) gives

$$G(x_1) = -\left(\frac{1}{2} - \frac{x_1}{z}\right) + W(\mu^S(x_1)) - W(0)$$

Observe that

$$G(0) = -\frac{1}{2} + W(\mu^S(0)) - W(0) < -\frac{1}{2} + \overline{W} - \underline{W} < 0$$

Evaluated at a_1, we have

$$G(\alpha_1) = -\left(\frac{1}{2} - \frac{\alpha_1}{z}\right) + W(\mu^S(a_1)) - W(0) > H(\alpha_1) = 0$$

The inequality passes because $\frac{\alpha_1}{z} > \alpha_1$ and $W(\mu^S(\alpha_1)) > W(\mu_1)$ (the agent's reputational boost from success after pooling is always less than the reputational boost of success following partial pooling).

Finally, notice that

$$G'(x_1) = \frac{1}{z} + W'(\mu^S)\frac{\partial \mu^S}{\partial x_1} > 0$$

This inequality follows because: (1) $W'(\mu^S) > 0$ and (2) $\frac{\partial \mu^S(x_1)}{\partial x_1} > 0$.[12] It is immediate that a number $\alpha_1 \in [0, a_1)$ solves $G(\alpha_1) = 0$

■

We establish the equilibrium in the following regions. Region (1) lies between $(\alpha_1, \frac{1}{2}]$; Region (2) lies between $[a_1, \alpha_1)$; Region (3) lies between $[0, a_1)$

Region (1) Perfect Pooling

In this interval of global facts, the agent perfect pools and the principal affirms all dispositions. To perfectly pool, the agent sets $y_1^* = 1 - x_1$. In that case, the principal's beliefs following success are:

$$\mu^S = \frac{\mu_1}{\mu_1 + (1 - \mu_1)} = \mu_1$$

Given perfect pooling, the principal always affirms the disposition. The agent's payoff to perfect pooling is:

$$-(1 - x_1 - \frac{1}{2}) + W(\mu_1)$$

[12] (note to lewis and scott, in doing this derivative remember to first do the replace $y^* = 1 - \frac{x_1}{z}$; otherwise the sign is off)

Given affirmance is in the offing, the agent's payoff to selecting a different cutline $y_1 < 1 - x$ is

$$-(y_1 - \frac{1}{2}) + W(0)$$

The agent's best possible deviation sets $y_1 = \frac{1}{2}$. This deviation is unprofitable if

$$-(1 - x_1 - \frac{1}{2}) + W(\mu_1) > W(0)$$

or

$$H(x_1) > 0$$

In the region $[\alpha_1, \frac{1}{2}]$, lemma 2 shows that this is always true. As such, the agent prefers to pool. Given perfect pooling, the principal affirms any disposition rather than suffer a loss with positive probability. The proof of the pooling region between $[\frac{1}{2}, \beta_1]$ is analogous.

Region (2) Partial Pooling Region-No Reversal $x_1 \in [a_1, \alpha_1)$

In this region, the agent selects the cutline such that he is indifferent between finding the claim valid and not invalid. The cutline is less than complete pooling, but the principal nonetheless affirms all dispositions. The principal affirms all dispositions in period 1 if

$$y_1^C > y^* = 1 - \frac{x_1}{z}$$

where y_1^C is the cutline selected by the biased agent. Assume that the all dispositions are affirmed (I will show this is true in a moment). In that case, the agent is indifferent between finding a claim valid and not when

$$Z(y_1^c(x_1), x_1) = -(y_1^c(x_1) - \frac{1}{2}) + W(\mu^s(y_1^c(x_1), x_1)) - W(0) = 0 \qquad (14)$$

where $y_1^c(x_1)$ solves equation (14). Note, further,

$$\frac{\partial Z}{\partial y_1^c} = -1 + W' \frac{\partial \mu^s}{\partial y} < 0$$

The expression decreases in y_1^c holding constant x_1.

The expression $G(x_1)$ is the same as $Z(y_1^c(x_1), x_1)$ when $y_1^c(a_1)$ equals $y^*(a_1)$.

Combined with the definition of a_1, it follows that

$$G(a_1) = Z(y_1^c(a_1), a_1) = Z(y^*(a_1), a_1) = 0$$

Recall that $G(x_1)$ increases in x_1. As a result, for all global fact values between $[a_1, \alpha_1]$,

$$G(x_1) \ = Z(y^*(x_1), x_1) > 0$$

Thus, for these global facts, $y^*(x_1)$ is not the optimal cutline. Further, since Z is a decreasing function, we know that the biased agent's cutline must be strictly greater than the one that makes the principal indifferent between reversing and not ($y_1^c(x) > y^*$). As a result, the principal prefers to affirm all dispositions in this region and the agent's cutline is given by expression (14).

For global facts on the interval between $[\beta_1, b_1]$ do the same analysis, but replace $y_1^c(x_1)$ with $1 - y_1^c(x_1)$

(c) Partial Pooling Region-Reversal $[0, a_1]$

The argument from part (b) above shows that $y_1^c(x) < y^*$ for global facts in this range. Reputation alone doesn't provide enough incentives. At the agent's cutline strategy ($\widehat{y}_1$), the agent must be indifferent between validating the claim and not, given the probability of reversal of a validity claim, $\widehat{\gamma}_v$. The agent's indifference expression is

$$-(\widehat{y}_1 - \frac{1}{2}) + W(\mu^s(\widehat{y}_1, x_1)) = -\widehat{\gamma}_v(\widehat{y}_1 - \frac{1}{2} + k) + W(0) \tag{15}$$

At the same time, the principal must be indifferent between reversing and affirming, given beliefs consistent with the strategy $\widehat{y}$. As a result, it must be that

$$x_1 - (1 - \widehat{y}_1)z = 0 \tag{16}$$

The equilibrium is defined as the joint solution to (15) and (16) along with interim beliefs $\mu = \frac{\mu_1 x}{\mu_1 x + (1-\mu_1)(1-y^*)}$. Solving equation (16) for $\widehat{y}_1$ yields

$$\widehat{y}_1 = 1 - \frac{x_1}{z} \tag{A1}$$

Replacing $\widehat{y}$ in (5) gives

$$-(\frac{1}{2} - \frac{x_1}{z}) + \widehat{\gamma}_v(\frac{1}{2} - \frac{x_1}{z} + k) + W(\mu^s(\widehat{y}, x_1)) - W(0) = 0$$

or

$$\widehat{\gamma}_v = \frac{(\frac{1}{2} - \frac{x_1}{z}) - [W(\mu^s(\widehat{y}, x_1)) - W(0)]}{(\frac{1}{2} - \frac{x_1}{z} + k)} = \frac{-G(x_1)}{(\frac{1}{2} - \frac{x_1}{z} + k)} \tag{A2}$$

We know that $G(x_1) < 0$ for all values of $x_1 \in [0, a_1]$. Equation (A2) thus provides a positive value. The equilibrium values γ_v^*, y^{**} are given by the solutions to (**??**) and (**??**)

The analysis for global facts in the interval $[b_1, 1]$ is the much the same, except

$$\widehat{y}_1 = \frac{1 - x_1}{z} \tag{A3}$$

and

$$\widehat{\gamma}_{nv} = \frac{-G(1 - x_1)}{(\frac{1}{2} - \frac{1-x_1}{z} + k)} \tag{A4}$$

.■

Proof of Proposition 3

[TBD]

Proof of Proposition 4

Proof: Only the biased agent ever fails. Thus, the principal's belief after failure is $\mu_2^F = 0$. The second period bounds are defined as:

$$\begin{aligned} a_2 &= \frac{\frac{1}{2}}{1+1} = \frac{1}{4} \\ b_2 &= 1 - \frac{1}{4} = \frac{3}{4} \end{aligned}$$

Take a case with a global fact in the interval, $[\alpha_1, \frac{1}{2}]$. The biased agent perfectly pools ($y_1^* = 1 - x_1$). The principal's belief following success is

$$\mu_2^S = \frac{\mu_1}{\mu_1 + (1 - \mu_1)(1 - (1 - x_1 - y_1^*))} = \mu_1$$

Take a case with a global fact in the interval, $[\frac{1}{2}, \beta_1]$. The biased agent perfectly pools ($y_1^* = 1 - x_1$). The principal's beliefs following success is

$$\mu_2^S = \frac{\mu_1}{\mu_1 + (1 - \mu_1)(1 - (y_1^* - 1 - x_1))} = \mu_1$$

Success in this interval doesn't change the belief and thus has no effect on the discretion bounds.

Take case with a global fact in the interval $[0, a_1]$. The principal's belief

following success is

$$\mu_2^S = \frac{\mu_1}{\mu_1 + (1-\mu_1)(1-[(1-x)-y_1^*])}$$

where

$$y_1^* = 1 - \frac{x_1}{z}$$

Plugging in, we see that

$$\mu_2^S = \frac{\mu_1}{\mu_1 + (1-\mu_1)(1-\frac{x_1(1-z)}{z})}$$

And, we have

$$\frac{\partial \mu_2^S}{\partial x_1} = \frac{\left(\mu_2^S\right)^2 (1-\mu_1)(1-z)}{\mu_1 z} > 0$$

The inequality follows because $z < 1$. Thus the principal's belief following success increases in the interval.

Take a case with in the interval $[a_1, \alpha_1]$. The principal's belief following success is

$$\mu_2^S = \frac{\mu_1}{\mu_1 + (1-\mu_1)(1-[(1-x_1)-y_1^c(x_1)])}$$

where $y_1^c(x)$ is implicitly defined as the solution to

$$G(y^c(x_1)) = -(y^c - \frac{1}{2}) + W(\mu_2^S(y^c(x_1), x_1)) - W(0) = 0$$

The derivative of the beliefs equals

$$\frac{\partial \mu_2^S}{\partial x_1} = -\frac{-\mu_1(1-\mu_1)(1+\frac{\partial y_1^c}{\partial x_1})}{[\mu_1 + (1-\mu_1)((1-(1-x_1)-y_1^c(x_1)]^2}$$

The sign of the derivative is the sign of $-\left[1 + \frac{\partial y_1^c}{\partial x_1}\right]$.

Implicit differentiation of $G(y^c(x))$ gives

$$-\frac{\partial y_1^c}{\partial x_1} + W'(\mu_2^S(y_1^c, x_1))\frac{\partial \mu_2^S}{\partial x_1} + W'(\mu_2^S(y_1^c, x_1))\frac{\partial \mu_2^S}{\partial y_1^c}\frac{\partial y_1^c}{\partial x_1} = 0$$

And as a result, we have

$$\frac{\partial y_1^c}{\partial x_1} = \frac{-W'(\mu_2^S(y^c, x_1))\frac{\partial \mu_2^S}{\partial x_1}}{W'(\mu_2^S(y^c, x_1)\frac{\partial \mu_2^S}{\partial y^c} - 1} < 0$$

Notice that $\frac{\partial \mu_2^S}{\partial x_1} = \frac{\partial \mu_2^S}{\partial y_1^c}$. Using this fact, we can bound the right hand side: it must be larger than -1. Formally, we know that

$$-1 = \frac{-W'(\mu_2^S(y^c, x_1))\frac{\partial \mu_2^S}{\partial x_1}}{W'(\mu_2^S(y^c, x_1)\frac{\partial \mu_2^S}{\partial y^c}} < \frac{-W'(\mu_2^S(y^c, x_1))\frac{\partial \mu_2^S}{\partial x_1}}{W'(\mu_2^S(y^c, x_1)\frac{\partial \mu_2^S}{\partial y^c} - 1} = \frac{\partial y_1^c}{\partial x_1}$$

Rearranging the inequality gives:

$$\frac{\partial y_1^c}{\partial x_1} + 1 > 0$$

And so, it follows that:

$$\frac{\partial \mu_2^S}{\partial x_1} < 0$$

in the interval $\left[\underline{x}_1, \underline{\underline{x}}_1\right]$.

We can do the same analysis on the upper interval. The value of success is "hill shaped" with a flat spot in the middle. It increases then decreases then it is flat then it increases and decreases again.■