Political Agency, Oversight, and Bias: The Instrumental Value of Politicized Policymaking

Ian R. Turner*

November 28, 2016

Abstract

We develop a theory of policymaking between a principal, an agent, and an overseer. The agent increases the overall quality of policy outcomes through costly capacity investments. Oversight impacts agent investment incentives, but only if the policy bias of the agent does as well. The principal chooses whether or not to authorize the agent to make policy on her behalf in light of oversight dynamics. We show that when oversight is not an effective means of political control, the principal never benefits from a biased agent relative to an ally. When oversight is effective, the principal almost always benefits from agent bias. Based on the dynamics of policymaking oversight the ally principle fails for instrumental reasons. The principal benefits from trading off agent bias for higher capacity investments. The results have implications for bureaucratic appointments and intra-agency organization.

Word count: 8106 (In-text); 1818 (Supporting information); 9924 (Total=in-text+SI)

*Assistant Professor, Department of Political Science, Texas A&M University, Email: irturner@tamu.edu.

The vast majority of public policy is developed and implemented by bureaucratic agencies whose authority to do so was delegated from a political principal.[1] Delegation, however, introduces the potential for political agency problems: the ability of agents to subvert the wishes of their principal(s) and pursue their own goals. First, agents to whom authority has been delegated may be biased and pursue policy outcomes that diverge from those preferred by the principal, thereby subverting the principal's wishes (Gailmard 2002). Second, the agent may invest insufficient effort toward the production of high quality policy outcomes, sometimes referred to as slack (Bueno de Mesquita and Stephenson 2007; Turner 2016). Both possibilities can lead to low quality policy outcomes from the principal's perspective. In short, bureaucracies are tasked with not only crafting the substance of policy, which allows for the introduction of bias, but also investing in capacity to effectively implement policy in practice, which introduces the potential for slacking.[2]

Two commonly proposed solutions to these agency problems are extensive oversight and delegating to allies. The logic underlying the ally principle is straightforward and has been shown to hold in diverse environments (Bendor and Meirowitz 2004).[3] All else equal, a principal prefers to authorize an ally agent because the agent can be expected to take actions in line with the principal's interests regardless of informational or experiential advantages (Gailmard and Patty 2013*a*, p.4-5). However, the situation becomes more complicated when the agent must also invest in capacity that improves the overall quality of outcomes. The principal, in this case, must also consider the provision of incentives to spur this investment, which can be supplied through oversight.

In the United States federal government perhaps the most famous example of this oversight

[1]The advantages of delegation include the ability to exploit the superior expertise contained in agencies (Epstein and O'Halloran 1994; Spence and Cross 2000), and the provision of incentives to specialize (Gilligan and Krehbiel 1987), gather information (Gailmard and Patty 2013*b*), and improve efficiency or reliability (Ting 2002, 2003). See Bendor, Glazer and Hammond (2001) for a comprehensive overview of theories of delegation, and Gailmard and Patty (2013*a*) for a treatment specific to bureaucratic politics.

[2]Of course, capacity within agencies can refer to many things including manpower, procedural development, technology to aid in processing, and the like. We use the term capacity as a catch-all term to capture the idea that being able to implement policy effectively — accurately process permits, conduct inspections, enforce compliance — requires investment in the ability to execute on the ground. Specifically, higher capacity leads to more precise implementation. This is closely related to what Carpenter (2001) calls "programmatic capacity," and can also be viewed through the lens of "street-level bureaucracy" (Lipsky 1980).

[3]For example, the principal prefers to choose a "clone" of herself as her agent (Gailmard and Patty 2013*a*) and grants more discretion to an ally agent (Epstein and O'Halloran 1994) when she wants to induce expertise acquisition (Gailmard and Patty 2007), or when uncertainty is high (Bendor, Glazer and Hammond 2001; Moe 2012).

is judicial review of bureaucratic agencies. Judicial review provisions are written into authorizing legislation when an agency is empowered to take action.[4] Essentially, the overseer is empowered to reverse (or, veto) agency policy actions. The principal then has not only authorized a bureaucratic agent to make policy but has also subjected the agent's actions to subsequent review, and possible rejection, by another political institution. Overseers such as courts also often have preferences that diverge from those of the principal, further compounding the political agency problems inherent in the policy process. This raises the question at the heart of this article: When does a political principal benefit from authorizing an agent to make policy on her behalf given that the agent, once authorized, makes policy in the shadow of oversight? Moreover, if the principal does benefit from delegating policymaking authority, what type of agent would she prefer?

In this paper we develop a political-institutional theory of policymaking and show that if both bias and slack are potential problems, the potential solutions — delegating to allies and extensive oversight — interact in unexpected ways. Specifically, we show that when oversight impacts agent capacity investment incentives the principal prefers to delegate to a biased agent rather than an ally (or "clone"). Oversight and agent bias interact in such a way that neither will strengthen investment incentives unless both do simultaneously. This has upstream effects on whether the principal benefits from delegation at all and, in situations in which she does, whether she is better off authorizing an agent that shares her preferences (an "ally") or a biased agent to make policy.

The ally principle is violated whenever oversight is effective at providing positive investment incentives for the agent due to the fact that the agent sets the substance of policy *and* invests in policy-improving capacity. Crucial to the theory is the fact that this insight only holds because of the intervening influence of oversight. Even if a principal benefits from delegating policymaking authority when oversight is ineffective, she never benefits from a biased agent relative to an ally. She would be better off with a "clone" agent in those situations. Otherwise, when oversight is effective, agent bias serves two interrelated purposes that increase the agent's capacity investments in equilibrium.

[4] Shipan (1997) provides a comprehensive study of the politics surrounding the choice of these provisions (see also McCann, Shipan and Wang 2016). Additionally, the Administrative Procedures Act (APA) directs courts to engage in so-called hard look review of agency regulations and overturn actions found to "arbitrary and capricious" (Breyer 1986).

First, it intensifies the agent's own motivations to acquire more capacity to ensure that his policies are realized. Second, it increases the stringency of oversight, which in turn increases the capacity the agent must develop in order to have his preferred policy realized. The principal can benefit from leveraging this dynamic by delegating policymaking authority to a biased agent. She can play the other actors' biases off one another to induce higher levels of capacity investment than would be possible if there were no effective oversight or if the principal engaged in oversight herself. Thus, while there are situations where the principal will delegate policymaking authority to an agent who shares her preferences, she is better off under particular institutional environments delegating that authority to a biased agent provided oversight is effective as an institutional check on agent behavior. It is instrumentally valuable for a political principal to delegate policymaking authority to an agent that will make policy in a politicized (i.e., ideologically contentious) policymaking environment.

We add to existing literature by shifting attention away from the role that oversight plays in constraining agents' substantive policy choices and toward how oversight structures agent incentives to invest in capacity that improves the overall implementation of policy. Previous work has focused heavily on oversight's impact on disciplining the substance of agent policy choices to be more closely aligned with their principal (e.g., Epstein and O'Halloran 1999; Patty and Turner 2016; Shipan 1997).[5] In contrast, our argument tracks the empirical reality that many oversight institutions like courts have moved heavily toward procedural, rather than substantive, review of administration actions in recent times (Kagan 2001). Specifically, courts are often most concerned with the way(s) in which agencies implement or apply policy, rather than the specific content of the policies being challenged.[6] This is to be expected as in many cases the overseer lacks the same

[5]More generally, previous work has highlighted how oversight impacts incentives for politicians to pander in electoral settings (Fox and Stephenson 2011) or to acquire information (Dragu and Board 2015), and induces more ideologically desirable policy (Wiseman 2009), as well as the invaluable insight provided by previous work examining signaling dynamics and their effect on substantive policy choice (see e.g., Boehmke, Gailmard and Patty 2006; Carpenter and Ting 2007; Gailmard and Patty 2007, 2013*a*,*b*,*c*; Gilligan and Krehbiel 1987, 1989; Gordon and Hafer 2005; Patty 2009; Stephenson 2006; Ting 2008). Finally, see Bueno de Mesquita and Stephenson (2007), Stephenson (2006), and Turner (2016) for recent work that highlights the ways in which oversight impacts effort incentives.

[6]For example, the Federal Emergency Management Agency (FEMA) was taken to court based on the processes in place to allocate housing assistance following Hurricanes Katrina and Rita (see *ACORN v. FEMA*, 463 F. Supp. 2d 26 (D.D.C. 2006)) and the Social Security Administration was similarly challenged in the 1960s and 70s due to the ways in which they terminated aid to families with dependent children (see *Goldberg v. Kelly*, 392 U.S. 254 (1970) and Derthick (1990), specifically 132-135). In the former case the agency had developed a computer program to process applications

substantive expertise that leads to agency problems between the principal and agent.

We develop a theory that holds agent bias fixed. We structure the model so that the agent invests in capacity to reduce outcome uncertainty prior to learning the nature of the policy environment. Then the agent learns about the environment and sets the (possibly biased) substance of policy spatially. The overseer, meanwhile, only reviews the agent's investment decision. The overseer reviews the agent to ensure he has sufficient capacity to produce high quality policy in practice. Substantively, this turns the focus to the effects of oversight on the incentives for agencies to develop effective procedures to more accurately reach permitting and licensing decisions, properly and effectively allocate public assistance and government benefits, conduct adequate inspections and improve enforcement, and generally implement policy well. This setup also implies that the agent in our model, if authorized to make policy, has full discretion to set the substance of policy, which depends on his bias in equilibrium. Overall, our analysis complements existing work focused on ameliorating the potential bias in the content of policy by directly addressing the relatively understudied effect of oversight institutions on the incentives for agents to develop capacity that leads to higher quality policy outcomes. Indeed, in our model bias can prove to be instrumentally valuable for the principal precisely because it can be coupled effectively with oversight institutions to strengthen agent investment incentives.[7]

This latter point also speaks to literature examining optimal agent bias. Bendor and Meirowitz (2004) show that principals may prefer biased agents to allies if biased agents are willing to work harder or have some type of beneficial valence that is correlated with their bias and benefits the principal. We provide a distinctly political-institutional rationale for why biased agents can be preferable

that led to erroneous housing assistance decisions. In the latter case the agency had failed to develop procedures allowing for individuals having public assistance benefits terminated to adequately appeal the agency's decision. In both cases, the courts ruled that the agencies' interests in keeping administrative costs low were not sufficient to excuse the existing (lack of) capacity the agencies had in place, and the right to due process was of central concern. Both observations highlight the importance of agencies developing capacity to implement policy effectively on the ground in order to ensure high quality implementation. Similarly, Huber (2007) provides a comprehensive study of the Occupational Health and Safety Administration (OSHA), part of which highlights areas where the lack of agency capacity for on site inspections led to low quality program implementation. All of these cases highlight instances in which lacking agency capacity to implement policy effectively had a negative impact on outcomes on the ground, which is the central focus of our argument.

[7]This insight also complements recent work on competitive policy development (Hirsch and Shotts 2015*a*,*b*) and some of the issues studied in Huber and McCarty (2004) and Ting (2011).

for principals concerned with motivating desirable investments in valence. The theory developed here does not assume that biased agents are *per se* better equipped to produce high quality policy. Rather, agent bias only motivates capacity investments if effective oversight is present in the policymaking environment. Our results perhaps most closely resemble studies that show that agent bias is useful in delegation environments to induce specialization (Gilligan and Krehbiel 1987, 1989) or generate bargaining power (Gailmard and Hammond 2011).[8]

For instance, Gailmard and Hammond (2011) argue that the House of Representatives creates biased committees to increase House bargaining power relative to the Senate. The authors write that, "an unrepresentative committee is a veto constraint for the other chamber..." (p. 541). In our theory, the principal benefits from a biased agent precisely because she is able to sidestep her own commitment problems by leveraging those of the overseer. A biased overseer represents a "tougher veto point" with respect to agent policymaking, which the agent responds to by investing in higher levels of capacity to satisfy the overseer. The more divergent agent and overseer ideal points are, within reasonable limits, the more this dynamic intensifies these incentives. While the logic between our theory and this body of work are related, we extend and complement it. First, as noted above, we incorporate both the substantive setting of policy and investments to improve outcomes in one framework. Both are key to our results. Second, in our theory the presence of effective oversight is a necessary condition for agent bias to strengthen investment incentives.

Overall, the theory developed in this paper provides novel insight into the institutional and policy environments in which biased policymaking can benefit political principals. When oversight is an effective tool for political control, if the principal benefits from delegation at all, she benefits from delegating authority to a biased agent who will subsequently face a biased overseer in the policymaking game. Without the intervening dynamics of effective oversight the ally principle holds.

[8]More generally, several previous studies have shown that, at times, principals prefer biased agents based on divergent beliefs (Che and Kartik 2009), the optimal distribution of tasks between agents and reviewers (Bubb and Warren 2014), the need to induce information disclosure (Dessein 2002), to incentivize costly investment in policy development (Hirsch and Shotts 2015*a*), and to reduce rent-seeking by electorally motivated politicians (Van Weelden 2013). We provide results that are similar in the sense that they also show why political principals may prefer a biased agent, however, we diverge from previous work by analyzing an environment in which the institution of oversight is a necessary condition for biased agency to be beneficial.

Otherwise political principals derive instrumental value from the dual usage of oversight and agent bias as institutional motivators when bias and slack are both concerns.

The model

We study a simplified model of policymaking between three players: a principal (P) that chooses whether to authorize an agent to make policy on her behalf, an agent (A) that, if authorized by the principal, makes policy, and an overseer or reviewer (R) that has the power to invalidate agent policy actions. The game tracks the formation of a policy, denoted by y. The final policy outcome depends on players' choices within the game. Agent-made policy is only realized if (1) the principal chooses to delegate policymaking authority to the agent, and (2) the overseer chooses to uphold (rather than overturn) the agent. If the principal does not delegate authority or the agent is granted policymaking authority but the overseer reverses him, then an unregulated outcome is realized. This represents an environment in which the principal does not have the capacity to implement policy herself, or, equivalently, it is prohibitively costly for her to do so. Players desire final policy to coincide with their preferences as closely as possible.

The nature of the policy environment is captured by a true state of the world, denoted by $\omega \in \Omega = \mathbb{R}$, that is drawn by a non-strategic player ("Nature") at the beginning of the game according to cumulative distribution function F, which is symmetric around mean 0 with strictly positive, finite variance V_F. The state variable ω is meant to capture the contingencies of the policy environment and its realization can therefore be understood as being predicated on the unregulated private interactions between individuals or firms. Conditional on players' preferences and actions, the realization of ω leads to the potential for biased policy output. Only the agent learns ω, which occurs only if the principal delegates authority (i.e., the agent is an "expert" relative to the principal and overseer). The other players never observe ω, but the characteristics of F are common knowledge. There is also an *implementation shock* that can shift agent-made policy away from its substantive intent. This captures the possibility for slack because the agent can reduce the variance of the shock with ex ante capacity investments. Let $\varepsilon \in \mathbb{R}$ represent this shock and be distributed according to cumulative distribution function $G_\varepsilon(e)$ that is symmetric around mean 0 with strictly positive, finite variance $V_\varepsilon(e)$. The likely magnitude of the shock is decreasing in agent capacity. Specifically, the variance of $G_\varepsilon(e)$ is continuously strictly decreasing and convex in agent investments (denoted by e). This ensures $V_\varepsilon(e) < V_\varepsilon(e')$ if and only if $e > e'$.[9] The more the agent invests in capacity the more precise is agent-made policy. The characteristics of $G_\varepsilon(e)$ are common knowledge, but no player observes ε directly. Put simply, the more capacity the agent acquires through his ex ante investment the more likely it is that outcomes will be realized at their intended spatial location.

Given this setup, both bias and slack can lead to inefficient agent-made policy outcomes. Insufficient ex ante capacity investments that improve policy can lead to inefficient outcomes *even when the principal and agent are ideological allies*. Similarly, agent bias can lead to the substantive content of policy being set away from ω, which harms principal welfare when principal-agent preferences diverge *even when the agent has invested maximally in capacity*. Thus, both bias and slack are omnipresent concerns for a principal interested in policy matching the true state in this model.

The timing of the game proceeds as follows. First, Nature (privately) draws ω according to F. The principal then, without observing ω, chooses whether to authorize the agent to make policy on her behalf. This choice is denoted by $a \in \{0,1\}$ where $a = 1$ represents a choice to delegate and $a = 0$ to not. If $a = 0$ then the game ends, ω obtains unencumbered by agent intervention, and payoffs are realized. If $a = 1$ then the agent chooses a capacity investment level $e \in [0,1]$. Following investment, the agent learns ω and chooses a substantive policy target $x \in X = \mathbb{R}$. The ex ante nature of the agent's investment focuses on realistic policymaking environments in which agencies must develop capacity to adequately implement policies once they are authorized to act.[10] Finally, the overseer observes the agent's capacity investment e, but not x, y, or ω, and chooses to either uphold or overturn.[11] This choice is denoted by $r \in \{0,1\}$ where $r = 0$ represents deference to the agent and

[9] That is, $\forall e > e'$, $G_\varepsilon(e)$ second-order stochastically dominates $G_\varepsilon(e')$.

[10] For instance, Carpenter (2001) distinguishes an agency's analytic and programmatic capacity. The former relates to technical knowledge, or knowledge of ω in our model, while the latter represents the agency's ability to effectively apply policy in practice, which we proxy through ex ante capacity investments. More generally, this set up tracks a growing literature on agency capacity (Huber and McCarty 2004; Ting 2011).

[11] Obviously, this is a strong informational assumption. We employ this assumption because it allows us to focus acutely on the effects of oversight on agent effort investment while holding bias fixed. Moreover, it allows us to substantively focus directly on the role of procedural review of agent actions on the incentives to produce high quality policy

$r = 1$ reversal. If $r = 1$ then the game ends, ω is realized, and players receive their payoffs. If $r = 0$ then Nature draws ε according to $G_\varepsilon(e)$, which is conditional on the agent's choice of e, the game ends, agent-made policy obtains, and payoffs are realized.[12] Accordingly, final policy outcomes are realized according to the following function,

$$y = \begin{cases} x - \omega + \varepsilon & \text{if } a = 1 \text{ and } r = 0, \\ -\omega & \text{if } a = 0 \text{ or } a = 1 \text{ and } r = 1. \end{cases} \tag{1}$$

Each players' induced preferences over policy depend on their respective "type" or ideal point, denoted by $t_i \in \mathbb{R}$, $i \in \{P, A, R\}$. Each players' ideal point dictates their welfare-maximizing policy outcome relative to ω. We normalize the principal's ideal point so that $t_P = 0$, which implies that the principal is solely concerned with final outcomes matching the state. We also assume that the overseer's ideal point is to the left of the principal so that $t_R < 0$. The analysis focuses on how oversight, agent authorization, and policymaking incentives vary as t_A varies relative to the other players' ideal points. The payoffs of the principal, the overseer, and the agent are given by the following expressions, respectively:

$$\begin{aligned} u_P(e,y,r) &= -y^2 - ca, \\ u_R(e,y,r) &= -(y - t_R)^2, \\ u_A(e,y,r) &= -\beta(y - t_A)^2 - \kappa e - \pi r. \end{aligned}$$

whatever the substantive content. As noted above, this assumption is stronger in appearance than in practice. In many important policy areas overseers such as courts focus directly on process or procedure including agencies' investments in detailed, high quality implementation plans (e.g., Melnick 1983, p. 55-57), the capacity agencies possess to enforce policy (e.g., Huber 2007), and the processes through which services are provided (or not provided) (e.g., Derthick 1990, p. 132-135). Still, what happens when this assumption is relaxed is an interesting, and important, question. While it is beyond the scope of this particular article see (Patty and Turner 2016) for a formal treatment of agent incentives when policy content is reviewed. One of the main results of their analysis is that the agent is induced to obfuscate with his policy choices when they are directly observable by an overseer. This also provides a reason to focus on effort in this article: this obfuscation leads to highly ambiguous effects on ex ante effort investments as modeled in this paper. Thus, we set aside fuller treatment of this alternative information structure for future work.

[12] The important part of Nature's choice of ε is that no players observe ε, and therefore cannot react to it, but all players understand how capacity investments impact the realization of ε given their common knowledge regarding $G_\varepsilon(e)$.

As noted above, the principal wants outcomes that match ω, but does take into account her potential authorization costs. If the principal chooses to authorize the agent she incurs authorization costs $c \geq 0$. This captures the fact that delegating authority to an agent requires an investment by the principal. Legislatures must write authorizing legislation and allocate budgetary resources, the President must outline administrative goals or directives, and agency heads must establish policy goals, staff departments, and outline procedures to direct the actions of bureaucratic subordinates. In all of these cases the principal incurs direct or indirect costs associated with making the choice to authorize an agent to make policy on her behalf. If the principal chooses not to authorize the agent to make policy then she forgoes paying this cost, but she must accept the realization of unregulated outcomes (ω). The overseer seeks to minimize the distance between realized policy (y) and its ideal point t_R. The agent also desires policy outcomes to be realized as close as possible to his ideal point, but his policy motivations, relative to the other components of his utility and the motivations of the other players, is captured by $\beta > 0$. Agent policy motivations increase in β. This can represent stronger "sense of mission" within an agency (Wilson 1989), a higher ratio of zealots to slackers (Gailmard and Patty 2007) or political appointees to career civil servants (Lewis 2008), or simply higher intrinsic policy motivations for the bureaucratic agent (Prendergast 2007). All else equal, players prefer more effective implementation generated through increased agent capacity, but only the agent bears the costs of that investment, denoted by $\kappa > 0$. This cost captures intuitive concepts of building bureaucratic capacity like increased staffing, investing time and resources toward streamlining procedures, or expanding enforcement programs (Huber 2007). Finally, the agent is also averse to being overturned by the overseer, captured by $\pi > 0$. The agent becomes more averse to being overturned as π increases. While we are agnostic as to the microfoundations of this parameter, it captures intuitive, realistic concepts based on considerations like career concerns, e.g., agent reputational losses for looking incompetent, budgetary considerations, etc. It suffices to simply think of π as a reversal cost the agent must internalize if he is overturned. The parameters are exogeneous and common knowledge.

We utilize perfect Bayesian equilibrium (PBE) in weakly undominated strategies. The principal's strategy consists of an agent authorization choice. Denote this strategy by s_P and the principal's equilibrium authorization choice by $a^* \in \{0,1\}$. The principal also has beliefs over ω and ε, which are represented by μ_P, a cumulative distribution function that represents a probability distribution over ω and ε. The agent's strategy consists of a capacity choice denoted by s_A^e, and a policy mapping conditional on the realization of ω denoted by $s_A^x(\omega)$. Further denote the agent's equilibrium capacity investment as e_A^* and his equilibrium substantive policy choice conditional on ω as $x_A^*(\omega)$. The agent also has beliefs over ε denoted by μ_A. The overseer's review strategy consists of a mapping from the set of agent capacity investment levels and the potential policy outcomes into a review decision. Denote this strategy by $s_R(e)$ that holds for any agent capacity investment $e \in [0,1]$ and potential policy outcome $y \in \mathbb{R}$. The overseer also has beliefs over ω and ε characterized in the same manner as the principal's beliefs, which are denoted by μ_R. A PBE is a complete profile of strategies and beliefs $\rho = (s_P, \mu_P, s_A^e, s_A^x, \mu_A, s_R, \mu_R)$ such that all players are maximizing their expected payoffs given other players' strategies and, when applicable, beliefs are consistent with Bayes' rule.[13]

Oversight, bias, and capacity investment

In this section we analyze the interactions between the agent and the overseer assuming that the principal has authorized the agent to make policy. To begin, the agent will always set policy at his ideal point. That is, the agent's equilibrium substantive policy choice is $x^*(\omega) = \omega + t_A$. This is a weakly dominant strategy for the agent, independent of his capacity investment and the overseer's oversight strategy, because the overseer does not observe x directly. This feature of the equilibrium can be thought of as the agent making *sincere* policy choices (from his point of view). It also isolates the effects of oversight on agent capacity investment incentives and the principal's potential ability to exploit agent bias to reduce slack. We now turn to the overseer's optimal oversight strategy.

The overseer's equilibrium strategy is driven by the desire to minimize the distance between its ideal point and realized outcomes. However, oversight is limited to a veto of agent-made policy. Courts, executive reviewers, and intra-agency veto points can often only accept or reject policies

[13]Given the set-up, these beliefs will always be pinned down by Bayes' rule.

rather than supplant them with their own policy (Bueno de Mesquita and Stephenson 2007). The overseer, upon observation of the agent's capacity e, can only accept the expected losses from upholding agent policy actions or overturn the agent and accept the expected losses from allowing unregulated outcomes to obtain. With this in mind, the overseer's net expected payoff from upholding the agent is given by,[14]

$$\Delta U_R(\text{uphold: } r=0;\rho_{-i}) = -t_A^2 + 2t_At_R - V_\varepsilon(e) + V_F.$$

Incentive compatibility implies that the overseer will uphold the agent, given his bias t_A and observed capacity investment e, if and only if $\Delta U_R(\text{uphold: } r=0;t_A,e) \geq 0$. Rearranging the net expected payoff yields the incentive compatibility condition for the overseer to uphold an agent with bias t_A who invested e in capacity:

$$\underbrace{V_F - V_\varepsilon(e)}_{\text{Precision improvement}} \quad \geq \quad \underbrace{t_A^2 - 2t_At_R}_{\text{Net spatial policy losses}} \tag{2}$$

Equation 2 provides an intuitive condition that must be met for the overseer to uphold the agent. The agent must invest enough in capacity to improve the precision of policy outcomes, relative to the volatility of the underlying policy environment, to offset any spatial policy losses incurred by his bias. The more capacity the agent develops to improve outcomes the more likely it is equation 2 will be satisfied. Conversely, the more biased the agent is relative to the overseer the less likely it will be satisfied. This implies that the more biased the agent is relative to the overseer, the more stringent is oversight. However, the more volatile the underlying policy environment, the less stringent is oversight. The agent making policy becomes more important the more volatile the underlying policy environment becomes. This highlights a commitment problem for the overseer: the more the agent is needed to regulate, the less demanding oversight is with respect to capacity investments.

[14]We use the notation $U_i(\cdot;\cdot)$, $i \in \{P,A,R\}$ to represent players' expected utility given their proposed action and those of the other players. We also use $\Delta U_i(a;\rho_{-i}) \equiv U_i(a;\rho_{-i}) - U_i(b;\rho_{-i})$ to represent the net expected payoff for player i taking action a instead of action b given the expected behavior of the other players, $-i$, in equilibrium.

Since the precision of policy outcomes is strictly increasing in agent capacity investments, the overseer's equilibrium strategy is equivalent to an investment threshold. Denote this threshold as $\underline{e}_R(t_A)$. If such an e exists, this threshold is the minimum level of capacity investment an agent must make in order to be upheld by the overseer given his bias: assuming such an e exists, $\underline{e}_R(t_A) \equiv e$ such that $V_F - V_\varepsilon(e) = t_A^2 - 2t_A t_R$. A capacity investment that solves equation 2 with equality may not always be feasible. In cases where there is no $e \in [0,1]$ that solves equation 2 with equality, the overseer either always overturns or always upholds the agent. We discuss these scenarios in greater detail below. When there does exist a feasible e that solves equation 2, the agent must invest, at a minimum, that level of e to receive deference. For the remainder of the analysis we focus on the more interesting cases in which the agent can invest in capacity to satisfy the overseer's threshold as defined. This yields the following equilibrium oversight strategy,

$$s_R^*(e) = \begin{cases} \text{uphold: } r = 0 & \text{if } e \geq \underline{e}_R(t_A), \\ \text{overturn: } r = 1 & \text{otherwise.} \end{cases} \tag{3}$$

The impact of oversight on agent capacity investments depends crucially on the agent's bias relative to the overseer. If the agent is too biased then the overseer will never uphold the agent, regardless of investment levels. In this case the overseer is perfectly skeptical of regulatory intervention. This environment is one in which even if the agent makes a maximal capacity investment, $e = 1$, to improve implementation quality, he cannot offset spatial policy losses.[15] That is, if t_A is sufficiently extreme relative to t_R so that equation 2 fails to hold even when $e = 1$ then the overseer always prefers unregulated outcomes. Note that the level of agent bias that is *too biased* is increasing in the volatility of unregulated outcomes, V_F. The more an agent is needed to improve policy outcomes, the more biased he can be before the overseer becomes perfectly skeptical.

In this case the agent responds by never making positive capacity investments. If an agent with this level of bias makes any positive capacity investment, given the overseer will overturn with

[15]Recall also that this case (*perfect skepticism*) and the one to follow (*perfect deference*) are the regimes that would result from the lack of a capacity investment e that would solve equation 2 with equality, as discussed above.

certainty, he incurs a net utility loss proportional to the cost of that investment κ. Thus, when facing a perfectly skeptical overseer, the agent never invests in capacity.

On the other extreme, if the agent is too moderate relative to the overseer he will never be overturned. In this environment we say that the overseer is perfectly deferential to the agent. This is the case any time spatial policy losses are offset even when the agent invests nothing in capacity. That is, if t_A is sufficiently close to t_R such that equation 2 holds even when $e = 0$ then the overseer can never commit to overturning the agent. The level of bias that can support perfect deference is increasing in unregulated outcome volatility, V_F. All else equal, the more volatile unregulated outcomes become, the less stringent oversight becomes and the harder it is for the overseer to commit to overturning a relatively moderate agent. This reveals a pathological limitation of oversight in this model: if the agent *is not biased enough* relative to the overseer then oversight plays no effective role in the provision of agent investment incentives.

It may seem intuitive that in response to perfect deference the agent again never invests in capacity since he will be upheld regardless. However, the agent is intrinsically motivated to improve outcomes. While oversight as an institution does not impact capacity investments in this case, the agent's own motivations do. Since the overseer will never overturn the agent, the agent makes capacity investments based solely on his own motivations. Denote this choice by,

$$\hat{e}_A(\beta, \kappa) \in \arg\max_{e} -\beta V_\varepsilon(e) - \kappa e. \tag{4}$$

When the overseer is perfectly deferential, the agent chooses a level of investment as if there were no oversight. In this case the agent's capacity investment is greater than the overseer's threshold level of acceptable capacity investment: $\hat{e}_A(\beta, \kappa) \geq \underline{e}_R(t_A)$. This follows from the fact that oversight is not stringent enough to bind the agent's investment decision. Intuitively, the agent's investment in this case is increasing in his implicit policy motivations, β, and decreasing in costs, κ.

The final, most interesting, environment is one in which the agent's capacity investment *is* affected by oversight. In this case the overseer employs conditional-deference. The agent is biased enough away from the overseer that the agent's unconstrained capacity investment $\hat{e}_A(\beta, \kappa)$ is not sufficient to satisfy the overseer's threshold $\underline{e}_R(t_A)$. That is, given the arrangement of t_A and t_R in this environment, the agent's capacity investment based on his own motivations is not enough to satisfy the overseer's threshold. However, there is a level of investment that would satisfy this threshold and lead to deference.

Accordingly, the agent responds by deciding if he is better off making the threshold capacity investment required to be upheld or making no capacity investment and being overturned.[16] Consider the agent's net expected payoff for a capacity investment sufficient to be upheld,

$$\Delta U_A(e \geq \underline{e}_R(t_A); r^*(e) = 0) = \beta(t_A^2 + V_F - V_\varepsilon(e)) - \kappa e + \pi.$$

Incentive compatibility implies that the agent will make a capacity investment sufficient to be upheld if $\Delta U_A(e \geq \underline{e}_R(t_A); r^*(e) = 0) \geq 0$. Solving the agent's incentive compatibility condition for e so that it holds with equality, and bounding the problem to ensure that a feasible solution always exists, yields the maximum level of capacity investment the agent is willing to make to be upheld when facing a conditional-deference overseer:[17]

$$e_A^{\max}(t_A) = \max\left[\min\left[\frac{\beta(t_A^2 + V_F - V_\varepsilon(e_A^{\max}(t_A))) + \pi}{\kappa}, 1\right], 0\right]. \tag{5}$$

If the maximum level of capacity investment the agent is willing to make to be upheld exceeds the threshold required by the overseer then the agent does so. Otherwise, if $e_A^{\max}(t_A) < \underline{e}_R(t_A)$ then the agent invests nothing in capacity and accepts being overturned. Thus, when facing conditional-deference oversight ($\underline{e}_R(t_A) > \hat{e}_A(\beta, \kappa)$) the agent will make a capacity investment exactly equal to the overseer's threshold if and only if $e_A^{\max}(t_A) \geq \underline{e}_R(t_A)$, and invest nothing otherwise.

Taken collectively the oversight/capacity investment combinations described above imply the following optimal capacity investment strategy for the agent,

[16] Note that if it is not incentive compatible for the agent to invest the threshold level to be upheld then he makes zero capacity investment because any positive investment that fails to the meet the threshold results in a net utility loss equal to the cost of that investment, as in the perfectly skeptical case.

[17] The assumptions made are to ensure that a solution exists within the specified range, $[0,1]$. Specifically, since the fraction derived from rearranging the agent's incentive compatibility condition to solve for e can dip below 0 or rise above 1 we bound the problem using max and min operators to rule out these possibilities.

$$
s_A^{e*} = \begin{cases} \hat{e}_A(\beta,\kappa) & \text{if } \hat{e}_A(\beta,\kappa) \geq \underline{e}_R(t_A), \\ \underline{e}_R(t_A) & \text{if } \hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A) \text{ and } e_A^{\max}(t_A) \geq \underline{e}_R(t_A), \\ 0 & \text{if } \hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A) \text{ and } e_A^{\max}(t_A) < \underline{e}_R(t_A), \end{cases} \tag{6}
$$

where $\underline{e}_R(t_A)$ is defined as e such that equation 2 holds with equality, $\hat{e}_A(\beta,\kappa)$ is implicitly defined by equation 4, and $e_A^{\max}(t_A)$ is implicitly defined by equation 5.

There are a few aspects of the agent's equilibrium capacity investment strategy worth noting further. First, notice that the presence of an overseer can induce higher levels of capacity investment from the agent than if there were no oversight. This is the second case of s_A^{e*} in which $\hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A)$ and $e_A^{\max}(t_A) \geq \underline{e}_R(t_A)$. Conversely, the overseer can also induce the agent to invest less than he would otherwise. This is the third case of s_A^{e*} in which $\hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A)$ and $e_A^{\max}(t_A) < \underline{e}_R(t_A)$. In this case the overseer provides a "deterrence effect" for the agent. Since capacity investments are costly, the agent is deterred from investing in capacity at all because the overseer will not allow outcomes to turn out worse than the reversion level of policy precision (V_F), which in this case is not *bad enough* to induce the agent to invest more.[18] Combining all of the analysis above yields the subgame equilibrium for agent and overseer interactions when the agent has been authorized to make policy, embodied in the following proposition.

Proposition 1. *Suppose the agent is authorized to make policy by the principal. Then a PBE of the agent-overseer subgame is characterized by the following collection of strategies,*

1. *The agent makes capacity investments according to s_A^{e*}, given by equation 6,*
2. *The agent always sets policy at his ideal point, $x^*(\omega) = \omega + t_A$,*
3. *The overseer makes review decisions according to $s_R^*(e)$, given by equation 3.*

Figure 1 provides a graphical example of the agent-overseer subgame equilibrium. When the agent's ideal point is sufficiently close to the overseer's, oversight does not bind the agent's

[18]This deterrence effect is qualitatively similar to the "bail out effect" provided by judicial review identified in previous theoretical work (e.g., Bueno de Mesquita and Stephenson 2007; Fox and Stephenson 2011; Turner 2016).

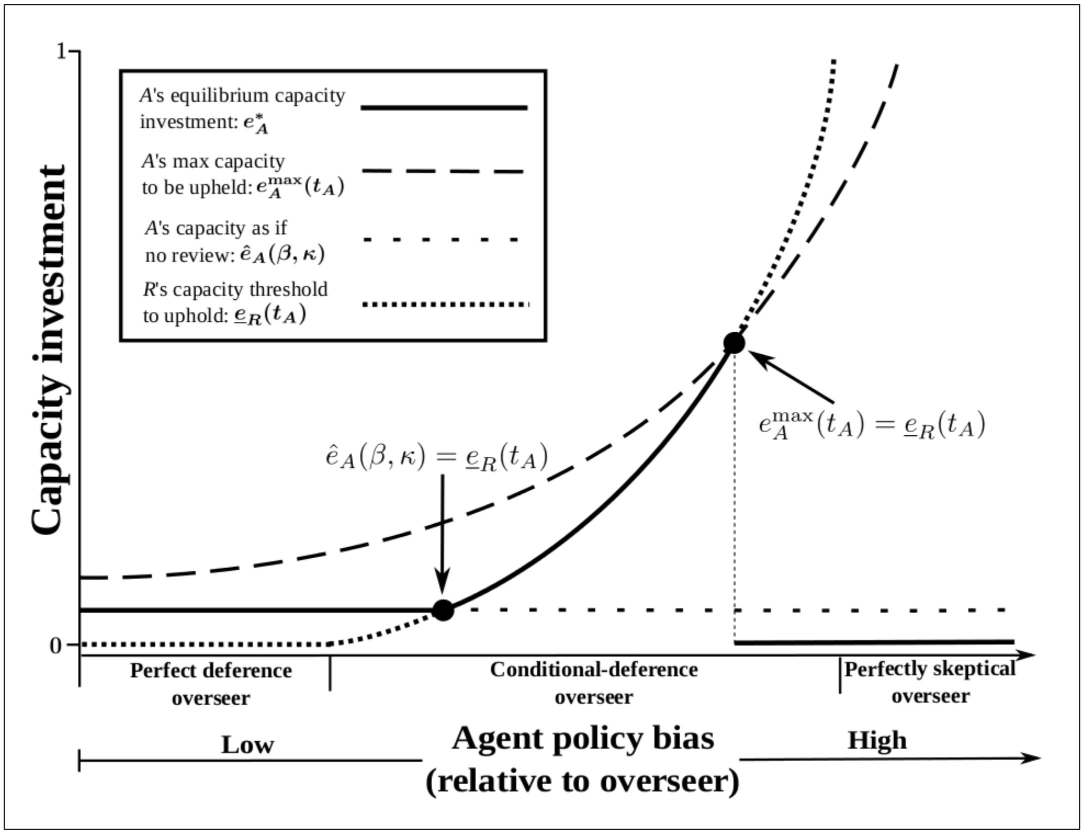


Figure 1: An Example of Equilibrium Capacity Investments Conditional on Agent Bias.
Note: The y-axis represents agent capacity investments, while the x-axis captures the distance between overseer and agent ideal points. Agent bias relative to the overseer is increasing left to right on the horizontal axis. The solid black line denotes the agent's equilibrium capacity investment given bias relative to the overseer, e_A^*. The black dashed line denotes the investment the agent is willing to make to be upheld given his bias, $e_A^{\max}(t_A)$. The wide-dotted (flat) line represents the agent's capacity investment when he will always be upheld, $\hat{e}_A(\beta,\kappa)$. The tight-dotted line denotes the overseer's threshold capacity investment required to uphold the agent, $\underline{e}_R(t_A)$.

capacity decisions. In this case, the agent invests in capacity based on his own motivations without taking into account the overseer since $\hat{e}_A(\beta,\kappa) > \underline{e}_R(t_A)$. As the agent becomes more biased there is a point at which the overseer begins to require positive capacity investments (the tight-dotted line begins increasing with agent bias). Past the point at which $\hat{e}_A(\beta,\kappa) = \underline{e}_R(t_A)$ the maximum capacity investment the agent is willing to make becomes the relevant consideration. So long as $e_A^{\max}(t_A) \geq \underline{e}_R(t_A)$ the agent invests at the overseer's threshold and is upheld. This is the intermediate region of equilibrium capacity investment in the figure where the solid black line tracks the overseer's threshold. As the agent continues to become more biased, oversight becomes more stringent and the agent becomes willing to invest more in capacity to be upheld. However, once the agent becomes too biased, i.e., $e_A^{\max}(t_A)$ drops below $\underline{e}_R(t_A)$, equilibrium capacity investments drop to 0. Oversight becomes too demanding and the agent is no longer willing to acquire enough capacity to be upheld. Figure 1 also provides visual intuition for the main result characterizing the relationship between oversight, bias, and agent capacity investments, described in the following proposition.

Proposition 2. *In equilibrium, agent bias strengthens agent capacity investment incentives if and only if oversight also strengthens these incentives.*

Proposition 2 presents a central result for the theory developed here. When oversight does not effectively strengthen investment incentives for the agent neither does agent bias. When the agent is not too biased he simply invests based on his own motivations (β and κ). Neither the agent's bias (t_A) nor oversight (through π) play a role in this investment. Similarly, when the agent is too biased, capacity investments are also invariant. They are always zero since the agent will always be overturned. In fact, in this case investment incentives are *weakened* and the agent is deterred from building any capacity. However, in the intermediate range of agent bias, capacity investments are increasing in both agent bias and the agent's aversion to being overturned, which only applies when oversight is effective. Thus, an agent's bias induces higher capacity investments if and only if the presence of oversight also does. This illustrates a fundamental interdependence between utilizing tools like appointing biased agents (or, "zealots") to direct agencies (Gailmard and Patty 2007) and institutionalized oversight to impact capacity investment incentives. One is not effective without the other. This raises two interrelated questions based on the dynamics described in this section. First, when does a political principal benefit from authorizing an agent to make policy on her behalf? Second, when delegation is beneficial does the principal benefit from authorizing an ally or biased agent to make policy?

The instrumental value of politicized policymaking

In this section we explore when, and under what circumstances, a political principal benefits from authorizing a biased agent to make policy on her behalf. The dynamics between capacity investments and oversight described in the previous section play a central role. If the principal chooses *not* to delegate authority she receives the following expected payoff,

$$U_P(a=0) = -V_F.$$

Since not authorizing the agent to make policy is equivalent to allowing unregulated outcomes to obtain, the principal loses utility equal to her expectation of these outcomes. Whether the principal finds it beneficial to authorize the agent depends on the relative locations of agent and overseer ideal points. If authorized, the agent will invest in capacity according to the subgame equilibrium characterized above. We analyze the principal's choices based on which environment the agent and overseer would interact within following delegation: perfectly skeptical, perfectly deferential, or conditional-deference.

The first case is one in which the agent will always be met with reversal. This is true when it is not incentive compatible for the agent to make capacity investments sufficient to be upheld, which could be because it is impossible to do so — the agent is so biased that the overseer will never uphold — or because oversight is too stringent and the agent is not willing to invest at the overseer's threshold — $e_A^{\max}(t_A) < \underline{e}_R(t_A)$. In either environment, if the principal does authorize the agent to make policy, knowing that he will ultimately be overturned, she can expect to simply bear the costs of authorization,

$$\Delta U_P(a=1; r^*(e^*)=1) = -c.$$

In terms of policy, this payoff is equivalent to the principal simply not authorizing the agent. In both instances, final outcomes are predicated on the unregulated actions of private individuals or firms. However, she must also incur the authorization costs to empower the agent. Intuitively whenever the principal would pay a positive cost, $c > 0$, for authorizing an agent to make policy just to see that agent's actions overturned, she is better off simply ending the game by not authorizing. When authorization is costless ($c = 0$) the principal is indifferent between authorizing the agent and not. Since costly authorization is more realistic we do not dwell on breaking this knife-edge case in a particular way.[19] Overall, when authorizing the agent does not impact policy outcomes and delegation is costly to the principal, she never authorizes the agent to make policy in equilbrium.

The second case is when the agent, if authorized to make policy, receives perfect deference. In this environment the agent develops capacity based on his own motivations. So the principal must decide if it is beneficial for her to allow the agent to make policy given that the agent will have unfettered discretion once authority is transferred. In this case the agent's actions will always obtain if he is authorized and the principal's corresponding incentive compatibility condition that must be met to authorize is given by,

$$\underbrace{t_A^2}_{\text{Spatial loss}} \leq \underbrace{V_F - V_\varepsilon(\hat{e}_A(\beta,\kappa))}_{\text{Precision improvement}} - \underbrace{c.}_{\text{Authorization cost}}$$

Intuitively, the principal benefits from authorizing the agent to make policy in this environment if the agent is not too biased. Specifically, the spatial losses associated with delegating authority to the agent must be outweighed by the improvement in policy precision induced given that the agent will always invest in capacity based on his own motivations, $\hat{e}_A(\beta,\kappa)$, less the costs of authorization. The likelihood this condition is met and the principal benefits from agent authorization is unambiguously decreasing in agent bias t_A since this has no bearing on the agent's equilibrium capacity investment. Further, because this investment level is invariant to agent bias, the likelihood that this condition will be met is increasing in the agent's intrinsic policy motivations β and the volatility of unregulated outcomes V_F, and decreasing in capacity and authorization costs, κ and c respectively.

Substantively, this highlights the fact that when oversight is ineffective at strengthening capacity investment incentives, the principal benefits from delegation based solely on agent and policy-environmental characteristics. If the agent is highly motivated, or if capacity costs are low, then it is more likely that delegation is beneficial. However, if the policy environment is relatively stable

[19]One can also imagine situations in which c could be negative. For instance, if the principal gains by "shifting blame" for policy failures to the agent then one could imagine that even though outcomes will not be appreciably different, the principal gains utility from being able to blame the agent for that failure and avoid external political costs like electoral challenge. The model could be easily extended to incorporate this possibility.

without agent policy intervention or the agent is extremely biased, perhaps through a process like agency capture, then it is unlikely that the principal benefits from delegation even with a formal institutional "check" like oversight in place.

The invariance of agent capacity investments in these two scenarios, in which oversight does not strengthen the agent's investment incentives, has clear implications for the instrumental value of authorizing a biased agent from the principal's perspective, captured by the following proposition.

Proposition 3. *If the agent will either always be overturned or always be upheld following delegation, the principal never benefits from a biased agent relative to an ally agent.*

In the environments in which the agent will always be overturned if he is authorized to make policy, agent bias has no impact on principal utility. Regardless of the agent's bias, policy outcomes are the same: unregulated, agent-free, outcomes will obtain. Therefore, the only thing that impacts the principal's payoff is the cost of authorization. In environments where the overseer is perfectly deferential, a biased agent strictly decreases the principal's utility. While there are cases where the principal benefits from authorizing a positively biased agent, it is strictly better for the principal if the agent is an ally. Since the agent's capacity investment is invariant to his bias, the principal can only lose utility from positive bias as it only increases the spatial policy losses associated with agent-made policy with no corresponding policy quality benefits, i.e., there is no associated increase in policy precision. Thus, even if delegating policymaking authority is desirable, the principal never benefits from a biased agent relative to an ally.

Now consider the most interesting case when the agent, if authorized, faces conditional-deference oversight. In this case, the environment is characterized by intermediately biased agents (relative to the overseer) in which equilibrium capacity investments are at the overseer's threshold ($e_A^{\max}(t_A) \geq \underline{e}_R(t_A)$). The agent, if authorized, will make capacity investments equal to $\underline{e}_R(t_A)$ and be upheld by the overseer. This, combined with the reversion utility of not authorizing the agent, implies the following net expected payoff for the principal given agent authorization,

$$\Delta U_P(a=1; r^*(\underline{e}_R(t_A))=0) = -t_A^2 - V_\varepsilon(\underline{e}_R(t_A)) - c + V_F.$$

Since the agent will be upheld for making capacity investments exactly at the overseer's threshold, the principal incurs the spatial policy losses associated with an agent with bias t_A, but gains as the agent's associated capacity investment improves implementation quality relative to the volatility of unregulated outcomes. Now, since we know from the agent-overseer subgame that the agent will make the overseer indifferent, we can reduce this net expected payoff by substituting the value of $V_\varepsilon(\underline{e}_R(t_A))$ when the overseer's incentive compatibility condition (equation 2) holds with equality. This reduces the principal's net payoff in this environment to,

$$\Delta U_P(a=1; r^*(\underline{e}_R(t_A))=0) \quad = \quad -2t_A t_R - c.$$

Incentive compatibility implies that the principal will authorize the agent to make policy if,

$$-2t_A t_R \geq c.$$

If authorization costs are positive, $c > 0$, then the principal can only benefit from empowering the agent to make policy if the agent and overseer are on opposite sides of her (i.e., t_A and t_R are oppositely signed). Since by assumption $t_R < 0$ this means that if the principal benefits from delegation at all then the agent is positively biased on the opposite side of the principal than the overseer: $t_A > 0$. If authorization is costless then it is possible that the principal can benefit from delgating to an ally agent. However, even in that case the principal's utility is increasing in agent bias. In this environment, the only time the principal benefits from delegating to an ally rather than a biased agent is when delegation is costless and agent and overseer ideal points are arranged such that if the agent were positively biased then the overseer would become perfectly skeptical—oversight would become too stringent—and the agent would respond by investing nothing in capacity and accept being overturned. Otherwise, the principal always prefers a positively biased agent. This leads to the result characterizing when politicized policymaking is instrumentally valuable to the principal.

Proposition 4. *If the agent will make capacity investments at the threshold level required by the overseer following delegation, and authorization costs are positive, then the principal benefits from* *a biased agent relative to an ally whenever she benefits from delegating policymaking authority.*

Propositions 3 and 4 provide the basis for the main theoretical insights of this paper. When oversight is ineffective at strengthening agent incentives agent bias is only detrimental to political principals. However, when oversight is effective at providing positive incentives for agent policy-making, the principal would prefer to have a biased agent to continue to strengthen these incentives. That is, agent capacity investment incentives are increasingly strengthened the more effective oversight is *and* the more biased the agent. In this way, the principal instrumentally prefers to trade off biased content of policy for increased capacity when oversight is an effective institutional check on agent behavior. By pitting a biased agent against a biased overseer, in particular having an agent oppositely biased from the overseer, the principal can benefit from the increased precision induced through agent capacity investments. This is true any time oversight has a positive impact on agent incentives and the principal would benefit from having an agent make policy at all.

Figure 2 provides a graphical representation of the intuition underlying propositions 3 and 4 using the example equilibrium capacity investments from figure 1. The axes are the same as in figure 1, but in this case only equilibrium capacity investments are graphed. When the agent will always be upheld his investments are unresponsive to increasing bias and therefore the principal is only harmed by increased agent bias; she prefers an ally agent. On the other extreme, when the agent is too biased relative to the overseer so that he will always be overturned, positive capacity investments never occur in equilibrium. In this case, the principal again derives no benefit from a biased agent. These two cases capture the principal's incentives within the environments of proposition 3. However, when the agent is intermediately biased relative to the overseer, he makes capacity investments that exactly match the overseer's threshold. In this case, equilibrium capacity is increasing in agent bias until the point at which he becomes too biased. If the principal benefits from delegating then she also benefits from positive agent bias relative to an ally, as in proposition 4. These theoretical insights provide several empirically relevant implications for bureaucratic politics.

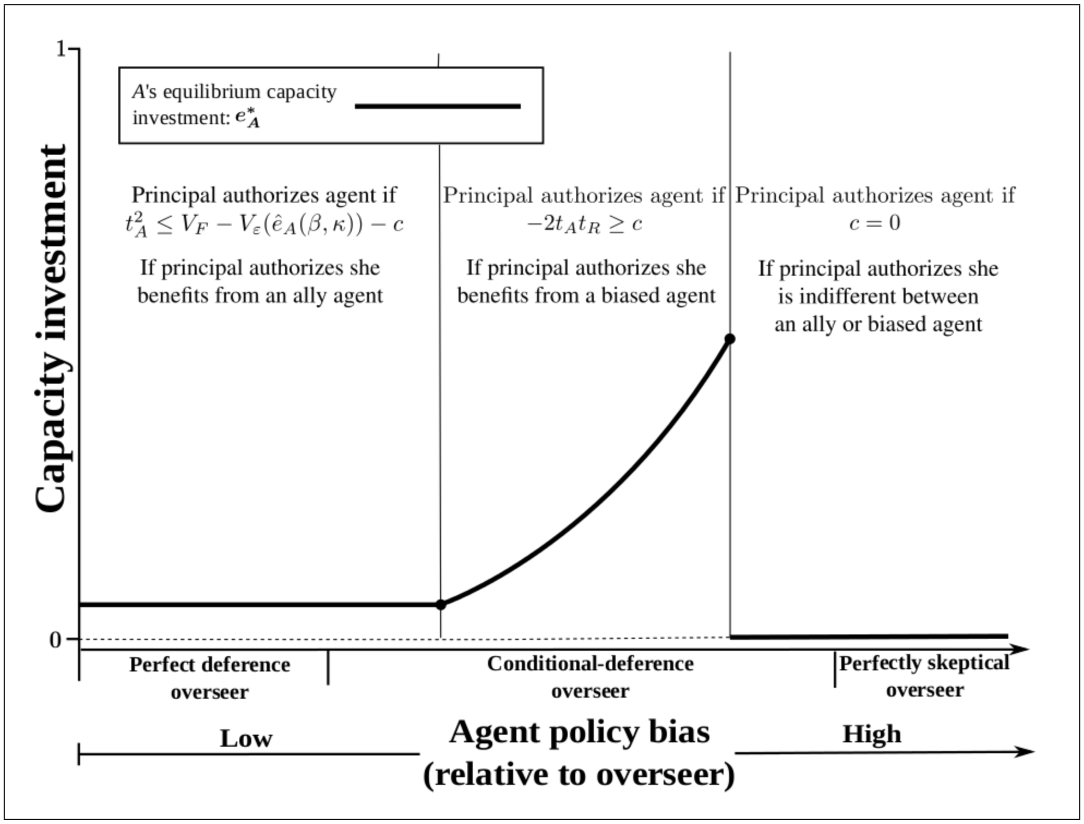
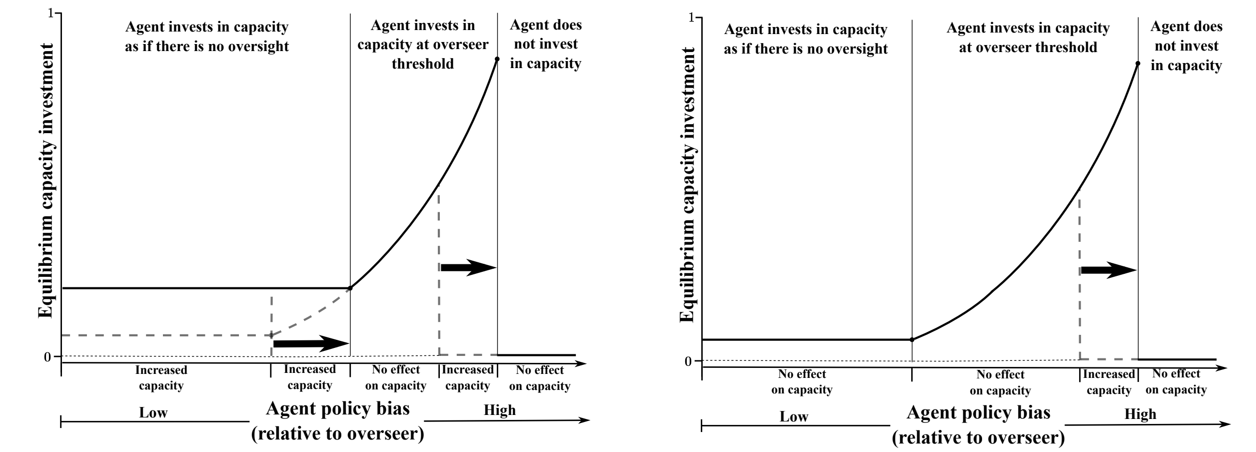


Figure 2: An Example of Equilibrium Authorization and the Instrumental Value of Agent Bias Conditional on Oversight Environment

Implications for bureaucratic politics

In this section we discuss the insights of the model for bureaucratic politics. First, we present two comparative statics — increased policy motivations β and increased reversal aversion π. Then, we proceed to discuss the implications for political appointments and intra-agency personnel politics.

First, consider a case in which agent policy motivations, β, increase, as illustrated in figure 3a. The effects of increasing motivations on incentives to invest in capacity depend on the nature of oversight. When the agent-overseer ideal points are relatively close to one another oversight has no effect on agent capacity investments. However, the agent's policy motivations do increase $\hat{e}_A(\beta,\kappa)$ and therefore, capacity investments increase proportional to the increase in β. This also expands the range of agent biases in which oversight does not impact these incentives. Once the agent becomes moderately biased, oversight is stringent enough to induce the agent to increase capacity investments



(a) Equilibrium Capacity for Increased β (b) Equilibrium Capacity for Increased π

Figure 3: Examples of Comparative Statics for Increased Policy Motivations and Reversal Aversion
Note: The gray dashed lines denote previous levels of equilibrium capacity investments prior to parameter increases. The black solid lines denote the equilibrium capacity investments following parameter increases. The arrows illustrate the increased range of agent biases following parameter shifts. The axes are as defined in Figures 1 and 2.

to be upheld. The increase in β, while it does increase the maximal investment the agent *is willing to make*, does not effectively alter equilibrium investment levels. By increasing policy motivations, however, the range of agents willing to invest enough to be upheld expands. This follows from the fact that increasing β increases $e_A^{\max}(t_A)$ without affecting $\underline{e}_R(t_A)$. More biased agents now find it beneficial to invest enough to avoid reversal. This further implies that the principal can benefit from a wider range of more extreme agent biases. Finally, extremely biased agents remain unaffected and are still unwilling to invest anything in capacity and accept being overturned. Thus, there is a positive correlation between agent policy motivations and agent bias in terms of principal gains from delegation, but this relationship is conditioned by the nature of oversight.

Now consider what happens when reversal aversion is increased, illustrated in figure 3b. When agents are ideologically proximate to the overseer, capacity investments remain unchanged since there is no risk of being overturned (i.e., π does not affect $\hat{e}_A(\beta,\kappa)$). The maximal investment the agent is willing to make to be upheld, $e_A^{\max}(t_A)$, does increase in π. This implies that the range of intermediately biased agents expands so that, as in the case of β, more extreme agents are now willing to invest sufficiently in capacity to receive deference. They switch from previously investing nothing and being overturned to investing at the required threshold and being upheld. Extremely biased agents remain unaffected by the change in π. Thus, strengthening an agent's aversion to reversal will increase observed equilibrium capacity investments, but only for a range of agent biases. From the principal's perspective agent bias and reversal aversion are complementary: an increase in π increases the maximal agent bias the principal can benefit from when delegating authority.

Taken together these comparative statics predict positive correlations between capacity investments and increased policy motivations and reversal aversion within agencies. However, these relationships are conditioned by the fact that these increases only 'work' for particular ranges of agent biases. In both cases the principal can benefit from a larger range of more extreme agent biases. Since the principal's utility is increasing in agent bias when oversight strengthens capacity investment incentives (the intermediately biased case), increasing agent policy motivations and reversal aversion increase the level of biases that benefit the principal. We now turn to applying these comparative statics to bureuacratic appointments and intra-agency organization.

One of the most important, and most difficult, tasks a president faces is staffing top positions in the federal bureaucracy (Waterman 1989). It is estimated that presidents must staff approximately 4,000 such positions upon taking office (Lewis 2008, 2011). The theory developed here, while appointments were not modeled explicitly, provides insight into what types of appointees can benefit presidents conditional on the nature of the larger institutional environment, e.g., the nature of oversight. Propositions 3 and 4 have clear implications for how presidents can leverage the institutional system in various ways to provide strong incentives for increased policy quality.

Lewis (2011) suggests that presidents benefit from appointing ideologically distinct agency heads when these appointees have difficulty affecting agency policy outputs in less ideologically friendly agencies (54-55). For example, suppose the EPA is largely staffed with pro-regulatory (e.g., "careerist") bureaucrats that seek to implement stringent environmental protection regulation, above and beyond what the President would prefer. It may be difficult for the EPA director to fully temper policy output and direct it back toward less stringent regulation. In this instance, our theory suggests that appointing an agency head as a "policy gatekeeper" that prefers less stringent regulation than the President will induce subordinate bureaucrats within the agency to produce higher quality regulatory interventions than if they were led by someone that shared their enthusiasm for stringent regulation. More generally, the results suggest that intra-agency conflict in the form of institutionalized gatekeepers or veto points can strongly incentivize bureaucrats to invest more into improving outcomes than they otherwise would in order to increase the probability that their policy goals are realized (Feldman 1989). The theory provides an instrumental rationale for bureaucratic organization that promotes a particular type of "internal conflict" in regulatory agencies (West 1988).

More generally, and in terms of the comparative statics described above, the president's incentives to authorize a biased agent, and the optimal level of that bias, is increasing as the agent becomes more policy motivated and reversal averse. That is, the more policy motivated and/or reversal averse is an appointee, the more the president benefits from the appointee's bias. In realistic environments in which oversight is much more difficult to alter and the overseer ideologically disagrees with the president, the president ultimately benefits from appointing someone who is oppositely biased from the overseer, that is highly policy motivated, and highly averse to being overturned by the overseer. As the latter characteristics increase, so does the level of bias that most greatly benefits the president. This effect is predicated on the fact that the appointee, once empowered to make policy on the presidents behalf, must take into account the overseer's preferences over policy and responds by investing in higher levels of capacity than would otherwise be obtained at lower levels of bias. That is, the more policy motivated or reversal averse the potential appointee, the more biased the appointee should be from a political principal's perspective when implementation incentives are of utmost importance.

Conclusion

In this paper, we developed a theory of delegation in the shadow of oversight and showed that political principals — presidents, legislatures, agency leaders — can benefit from authorizing biased agents to make policy on their behalf. This potential benefit is due to the recognition that policymakers both craft the substance of policy and invest in capacity to ensure those policies are implemented more effectively. Due to this duality in policymaking the principal benefits from pitting a biased agent, with full policymaking discretion, against an oppositely biased overseer, empowered to reverse the agent's actions if insufficient capacity is developed to improve the overall quality of outcomes. Institutionalized oversight is only effective as a means for strengthening incentives if the agent is biased, and leveraging agent bias to induce higher capacity is only a viable route to improve outcomes if oversight is an effective means of political control. The characteristics of the agent, the policy environment, and the dynamics of political oversight introduce both opportunities and constraints for principals interested in promoting strong incentives for agents they will authorize to make policy on their behalf. The model is flexible enough to be extended to include other important determinants of output such as interest group participation, oversight by multiple institutions, and allocation of policymaking tasks across multiple agents. This paper represents a step toward a fuller understanding of how ubiquitous processes, like bureaucratic policymaking in the shadow of oversight, impact the dynamics of political decisions like bureaucratic appointments and agency design.

References

Bendor, Jonathan and Adam Meirowitz. 2004. "Spatial Models of Delegation." *American Political Science Review* 98(2):293–310.

Bendor, Jonathan, Amihai Glazer and Thomas Hammond. 2001. "Theories of Delegation." *Annual Review of Political Science* 4(1):235–269.

Boehmke, Frederick, Sean Gailmard and John Patty. 2006. "Whose Ear to Bend? Information Sources and Venue Choice in Policy Making." *Quarterly Journal of Political Science* 1(2):139–169.

Breyer, Stephen. 1986. "Judicial Review of Questions of Law and Policy." *Administrative Law Review* 38(Fall):363–398.

Bubb, Ryan and Patrick L. Warren. 2014. "Optimal Agency Bias and Regulatory Review." *Journal of Legal Studies* 43(January):95–135.

Bueno de Mesquita, Ethan and Matthew C. Stephenson. 2007. "Regulatory Quality under Imperfect Oversight." *American Political Science Review* 101(3):605–620.

Carpenter, Daniel and Michael M. Ting. 2007. "Regulatory Errors with Endogenous Agendas." *American Journal of Political Science* 51(4):835–852.

Carpenter, Daniel P. 2001. *The Forging of Bureacuratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton, NJ: Princeton University Press.

Che, Yeon-Koo and Navin Kartik. 2009. "Opinions as Incentives." *Journal of Political Economy* 117(5):815–860.

Derthick, Martha. 1990. *Agency Under Stress: The Social Security Administration in American Government*. Washington, D.C.: The Brookings Institution.

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

Dragu, Tiberiu and Oliver Board. 2015. "On Judicial Review in a Separation of Powers System." *Political Science Research and Methods* 3(3):473–492.

Epstein, David and Sharyn O'Halloran. 1994. "Administrative Procedures, Information, and Agency Discretion: Slack vs. Flexibility." *American Journal of Political Science* 38(3):697–722.

Epstein, David and Sharyn O'Halloran. 1999. *Delegating Powers: A Transaction Cost Politics Approach to Policy Making Under Separate Powers*. New York, NY: Cambridge University Press.

Feldman, Martha. 1989. *Order without Design: Information Production and Policymaking*. Stanford, CA: Stanford University Press.

Fox, Justin and Matthew C. Stephenson. 2011. "Judicial Review as a Response to Political Posturing." *American Political Science Review* 105(2):397–414.

Gailmard, Sean. 2002. "Expertise, Subversion, and Bureaucratic Discretion." *Journal of Law, Economics, & Organization* 18(2):536–555.

Gailmard, Sean and John W. Patty. 2007. "Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise." *American Journal of Political Science* 51(4):873–889.

Gailmard, Sean and John W. Patty. 2013*a*. "Formal Models of Bureaucracy." *Annual Review of Political Science* 15:353–377.

Gailmard, Sean and John W. Patty. 2013*b*. *Learning While Governing: Expertise and Accountability in the Executive Branch*. Chicago, IL: University of Chicago Press.

Gailmard, Sean and John W. Patty. 2013*c*. "Stovepiping." *Journal of Theoretical Politics* 25(3):388–411.

Gailmard, Sean and Thomas Hammond. 2011. "Intercameral Bargaining and Intracameral Organization in Legislatures." *Journal of Politics* 73(2):535–546.

Gilligan, Thomas and Keith Krehbiel. 1987. "Collective Decision-Making and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics, and Organization* 3(2):287–335.

Gilligan, Thomas W. and Keith Krehbiel. 1989. "Asymmetric Information and Legislative Rules with a Heterogeneous Committee." *American Journal of Political Science* 33(2):459–490.

Gordon, Sanford C. and Catherine Hafer. 2005. "Flexing Muscle: Corporate Political Expenditures as Signals to the Bureaucracy." *American Political Science Review* 99(2):245–261.

Hirsch, Alexander V. and Kenneth W. Shotts. 2015*a*. "Competitive Policy Development." *American Economic Review* 105(4):1646–1664.

Hirsch, Alexander V. and Kenneth W. Shotts. 2015*b*. "Policy-Development Monopolies: Adverse Consequences and Institutional Responses." *Unpublished manuscript. California Institute of Technology* .

Huber, Gregory A. 2007. *The Craft of Bureaucratic Neutrality: Interests and Influence in Government Regulation of Occupational Safety*. New York, NY: Canbridge University Press.

Huber, John D. and Nolan McCarty. 2004. "Bureaucratic Capacity, Delegation, and Political Reform." *American Political Science Review* 98(3):481–494.

Kagan, Elena. 2001. "Presidential Administration." *Harvard Law Review* 114(8):2245–2385.

Lewis, David E. 2008. *The Politics of Presidential Appointments: Political Control and Bureaucratic Performance*. Princeton, NJ: Princeton Univesity Press.

Lewis, David E. 2011. "Presidential Appointments and Personnel." *Annual Review of Political Science* 14:47–66.

Lipsky, Michael. 1980. *Street-Level Bureaucracy*. New York, NY: Russell Sage Foundation.

McCann, Pamela J., Charles R. Shipan and Yuhua Wang. 2016. "Congress and Judicial Review of Agency Actions." *Working Paper. University of Southern California* .

Melnick, R. Shep. 1983. *Regulation and the Courts: The Case of The Clean Air Act*. Washington, D.C.: The Brookings Institution.

Moe, Terry. 2012. "Delegation, Control, and the Study of Public Bureaucracy." *The Forum* 10(2).

Patty, John W. 2009. "The Politics of Biased Information." *Journal of Politics* 71(2):385–397.

Patty, John W. and Ian R. Turner. 2016. "Ex Post Review and Expert Policymaking: When Does Oversight Reduce Accountability?" *Unpublished Manuscript. Texas A&M University. Presented at the 2016 Annual Meeting of the American Political Science Association* .

Prendergast, Canice. 2007. "The Motivation and Bias of Bureaucrats." *American Economic Review* 97(1):180–196.

Shipan, Charles. 1997. *Designing Judicial Review*. Ann Arbor, MI: University of Michigan Press.

Spence, David B. and Frank Cross. 2000. "A Public Choice Case for the Administrative State." *Georgetown Law Journal* 89(2000-2001):97–142.

Stephenson, Matthew C. 2006. "A Costly Signaling Theory of "Hard Look" Judicial Review." *Administrative Law Review* 58(4):753–814.

Ting, Michael M. 2002. "A Theory of Jurisdictional Assignments in Bureaucracies." *American Journal of Political Science* 46(2):pp. 364–378.

Ting, Michael M. 2003. "A Strategic Theory of Bureaucratic Redundancy." *American Journal of Political Science* 47(2):pp. 274–292.

Ting, Michael M. 2008. "Whistleblowing." *American Political Science Review* 102(2):249–267.

Ting, Michael M. 2011. "Organizational Capacity." *Journal of Law, Economics, & Organization* 27(2):245–271.

Turner, Ian R. 2016. "Working Smart *and* Hard? Agency Effort, Judicial Review, and Policy Precision." *Journal of Theoretical Politics* Forthcoming.
URL: *http://jtp.sagepub.com/content/early/2016/03/21/0951629816630430*

Van Weelden, Richard. 2013. "Candidates, Credibility, and Re-election Incentives." *Review of Economic Studies* 80(4):1622–1651.

Waterman, Richard W. 1989. *Presidential Influence and the Administrative State*. Knoxville, TN: University of Tennessee Press.

West, William F. 1988. "The Growth of Internal Conflict in Administrative Regulation." *Public Administration Review* 48(4):773–782.

Wilson, James Q. 1989. *Bureaucracy: What Government Agencies Do and Why They Do It*. New York, NY: Basic Books.

Wiseman, Alan E. 2009. "Delegation and Positive-Sum Bureaucracies." *Journal of Politics* 71(3):998–1014.

Supporting information

Agent-overseer subgame

Agent substantive policy choice.

Lemma 1. *The agent always sets the substantive content of policy at his ideal point:* $x^*(\omega) = \omega + t_A$.

Proof of Lemma 1. To show that the agent always sets policy at his ideal point we show that he is always weakly better off doing so by checking deviations in two cases: (1) when the overseer upholds the agent and (2) when the overseer reverses the agent. In both cases let $\delta > 0$ denote the agent's deviation so that if he deviates $x = \omega + t_A + \delta$.

Case 1: Overseer upholds. The agent's expected utility from setting $x = \omega + t_A$ is given by,

$$U_A(x = \omega + t_A | r = 0) = -\beta V_\varepsilon(e) - \kappa e.$$

The agent's expected utility from deviating to $x = \omega + t_A + \delta$ is given by,

$$U_A(x = \omega + t_A + \delta | r = 0) = -\beta(\delta^2 + V_\varepsilon(e)) - \kappa e.$$

These combine to give the agent's net expected payoff from deviating:

$$\Delta U_A(x = \omega + t_A + \delta | r = 0) \quad = \quad -\beta\delta^2.$$

Since $\beta > 0$ and $\delta > 0$ the agent is strictly worse off from deviating.

Case 2: Overseer reverses. When the agent is reversed the outcome does not vary regardless of the agent's choice of x. Thus, the agent is indifferent between setting policy faithfully at his ideal point and deviating to another policy. In both cases the agent's expected payoff is the same.

Taken together these two cases imply that the agent weakly prefers setting policy at his ideal point, as stated in the result. ■

Optimal oversight.

Lemma 2. *In equilibrium, the overseer plays the following best response strategy,*

$$s_R^*(e) = \begin{cases} \textit{uphold: } r = 0 & \textit{if } e \geq \underline{e}_R(t_A), \\ \textit{reverse: } r = 1 & \textit{otherwise,} \end{cases}$$

Proof of Lemma 2. First, consider the overseer's subjective expected utility for overturning the agent,

$$\begin{aligned} U_R(r = 1; \rho_{-R}) &= -(y - t_R)^2 \\ &= -\mathbb{E}[\omega - t_R]^2 - V[\omega], \\ &= -t_R^2 - V_F. \end{aligned}$$

Now, consider the overseer's subjective expected utility for upholding the agent,

$$\begin{aligned} U_R(r = 0; \rho_{-R}) &= -(y - t_R)^2, \\ &= -(x^*(\omega) - \omega + \varepsilon - t_R)^2, \\ &= -\mathbb{E}[x^*(\omega) - \omega - t_R]^2 - V[x^*(\omega) - \omega - t_R] - \mathbb{E}[\varepsilon|e]^2 - V[\varepsilon|e], \\ &= -(t_A - t_R)^2 - V_\varepsilon(e). \end{aligned}$$

Define $\Delta U_R(r = 0; \rho_{-R}) \equiv U_R(r = 0; \rho_{-R}) - U_R(r = 1; \rho_{-R})$ as the overseer's net expected utility for upholding. Then we have,

$$\begin{aligned} \Delta U_R(r = 0; \rho_{-R}) &= -(t_A - t_R)^2 - V_\varepsilon(e) + t_R^2 + V_F, \\ &= -t_A^2 + 2t_A t_R - V_\varepsilon(e) + V_F. \end{aligned}$$

Incentive compatibility implies that the overseer will uphold if and only if $\Delta U_R(r = 0; \rho_{-R}) \geq 0$.

Thus we have,

$$-t_A^2 + 2t_A t_R - V_\varepsilon(e) + V_F \quad \geq \quad 0.$$

Rearranging we have:

$$V_F - V_\varepsilon(e) \quad \geq \quad t_A^2 - 2t_A t_R, \tag{7}$$

as is presented in-text in equation 2. The increase in policy precision on the LHS must outweigh the net spatial policy losses based on divergent ideal points on the RHS. Now, by incentive compatibility the overseer's threshold level of required capacity investment to uphold the agent is defined as $\underline{e}_R(t_A) \equiv e$ such that equation 7 holds with equality given agent bias t_A, assuming such an e exists. ■

Agent capacity investments.

Lemma 3. *Define* $e_A^{\max}(t_A) = \max\left[\min\left[\frac{\beta(t_A^2+V_F-V_\varepsilon(e_A^{\max}(t_A)))+\pi}{\kappa}, 1\right], 0\right]$. *The agent will never make capacity investments higher than* $e_A^{\max}(t_A)$ *to be upheld by the overseer.*

Proof of Lemma 3. When the agent faces a conditional-deference overseer his net expected utility from investing in capacity at the threshold level required to be upheld is given by,

$$\Delta U_A(e \geq \underline{e}_R(t_A); \rho_{-A}) = \beta(t_A^2 + V_F - V_\varepsilon(e)) - \kappa e + \pi.$$

Thus, the agent will invest this level if and only if $\Delta U_A(e \geq \underline{e}_R(t_A); \rho_{-A}) \geq 0$. Solving the expression with equality for e gives the maximum level of capacity investment the agent would be willing to make given t_A in order to be upheld (by incentive compatibility):

$$e = \frac{\beta(t_A^2 + V_F - V_\varepsilon(e)) + \pi}{\kappa}. \tag{8}$$

The RHS of Equation 8 can fall below 0 and rise above 1. So to ensure a capacity investment always exists further define:

$$e_A^{\max}(t_A) = \max\left[\min\left[\frac{\beta(t_A^2 + V_F - V_\varepsilon(e_A^{\max}(t_A))) + \pi}{\kappa}, 1\right], 0\right].$$

Given this formulation, $e_A^{\max}(t_A)$ always exists. The RHS of equation 8 is continuous over the interval $[0,1]$. So, either $e_A^{\max}(t_A)$ is on a boundary or there is an interior solution, which is implied by (continuity and) the Intermediate Value Theorem. ■

Lemma 4. *In equilibrium, the agent makes capacity investments according to the following strategy,*

$$s_A^{e*} = \begin{cases} \hat{e}_A(\beta,\kappa) & \text{if } \hat{e}_A(\beta,\kappa) \geq \underline{e}_R(t_A), \\ \underline{e}_R(t_A) & \text{if } \hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A) \text{ and } e_A^{\max}(t_A) \geq \underline{e}_R(t_A), \\ 0 & \text{if } \hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A) \text{ and } e_A^{\max}(t_A) < \underline{e}_R(t_A), \end{cases}$$

where $\hat{e}_A = \arg\max_e -\beta V_\varepsilon(e) - \kappa e$, $\underline{e}_R(t_A) \equiv e$ *such that* $V_F - V_\varepsilon(e) = (t_A - t_R)^2 - t_R^2$, *and* $e_A^{\max}(t_A) = \max\left[\min\left[\frac{\beta(t_A^2 + V_F - V_\varepsilon(e_A^{\max}(t_A))) + \pi}{\kappa}, 1\right], 0\right]$.

Proof of Lemma 4. To verify that these are best responses for the agent we need to check three cases: (1) the overseer always upholds ($\hat{e}_A(\beta,\kappa) \geq \underline{e}_R(t_A)$); (2) the overseer always overturns ($\hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A)$ and $e_A^{\max}(t_A) < \underline{e}_R(t_A)$); (3) the overseer upholds if and only if the agent makes a large enough capacity investment, which is higher than the agent would invest absent oversight ($\hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A)$ and $e_A^{\max}(t_A) \geq \underline{e}_R(t_A)$). These cases are defined by the overseer's best response in Lemma 2 and the maximum capacity investment the agent is willing to make to be upheld in Lemma 3.

Overseer always upholds (perfectly deferential). The agent's expected payoff given he will be upheld is given by,

$$\begin{aligned} U_A(e|r=0) &= -\beta(\mathbb{E}[\varepsilon]^2 + V_\varepsilon(e)) - \kappa e, \\ &= -\beta V_\varepsilon(e) - \kappa e. \end{aligned}$$

The agent seeks to maximize $U_A(e|r=0)$ with his choice of e, which implies the following capacity investment,

$$\hat{e}_A(\beta,\kappa) \in \arg\max_{e} -\beta V_\varepsilon(e) - \kappa e.$$

Moreover, $\hat{e}_A(\beta,\kappa)$ exists since it is the maximum of a continuous function on a compact set and is unique so long as $V_\varepsilon(e)$ is strictly monotone.

Overseer always overturns (perfectly skeptical). To see why the agent never makes positive capacity investments in an environment in which he will always be reversed note that the agent's expected payoff for making positive capacity investments given he will be overturned is:

$$U_A(e>0|r=1) = -\beta(t_A^2 + V_F) - \kappa e - \pi.$$

The agent's expected payoff from investing nothing given he will be overturned is:

$$U_A(e=0|r=1) = -\beta(t_A^2 + V_F) - \pi.$$

These combine to give the agent's net expected payoff from making positive capacity investments given that he will be reversed by the overseer,

$$\begin{aligned} \Delta U_A(e>0|r=1) &= -\beta(t_A^2+V_F) + \beta(t_A^2+V_F) - \kappa e - \pi + \pi, \\ &= -\kappa e. \end{aligned}$$

Thus, if the agent makes positive capacity investments when he will be reversed he simply pays the cost for that investment, and, therefore, optimally invests nothing.

Conditional-deference overseer. In this environment $\hat{e}_A(\beta,\kappa) < \underline{e}_R(t_A)$ so the agent is constrained by the overseer. The agent compares his expected utility from investing in capacity at the threshold level and being upheld by the overseer and his expected utility from investing nothing ($e=0$) and being overturned. These expected payoffs are given by the following expressions, respectively:

$$\begin{aligned} U_A(e = \underline{e}_R; \rho_{-A}) &= -\beta V_\varepsilon(\underline{e}_R) - \kappa \underline{e}_R, \\ U_A(e = 0; \rho_{-A}) &= -\beta(t_A^2 + V_F) - \pi. \end{aligned}$$

These combine to give the net expected payoff for making capacity investments at the threshold (and being upheld rather than overturned):

$$\begin{aligned} \Delta U_A(\underline{e}_R; \rho_{-A}) &= -\beta V_\varepsilon(\underline{e}_R) - \kappa \underline{e}_R + \beta(t_A^2 + V_F) + \pi, \\ &= \beta(t_A^2 + V_F - V_\varepsilon(\underline{e}_R)) - \kappa \underline{e}_R + \pi. \end{aligned} \tag{9}$$

Equation 9 gives the agent's incentive compatibility condition for investing the threshold level, $\underline{e}_R(t_A)$, rather than $e = 0$ and being overturned. As long as this condition is weakly greater than zero the agent, in weakly undominated strategies, will make capacity investments at the threshold level required to be upheld when constrained by the overseer. ■

Proposition 1. *Suppose the agent is authorized to make policy by the principal. Then a perfect Bayesian equilibrium of the agent-overseer subgame is characterized by the following collection of strategies,*

1. *The agent makes capacity investments according to s_A^{e*}, given by equation 6,*
2. *The agent always sets policy at his ideal point, $x^*(\omega) = \omega + t_A$,*
3. *The overseer makes review decisions according to $s_R^*(e)$, given by equation 3,*

Proof of Proposition 1. This follows from a straightforward combination of lemmas 1, 2, 3, and 4. Lemmas 3 and 4 yield number 1 in the proposition, lemma 1 yields number 2, and lemma 2 yields number 3. ■

Proposition 2. *In equilibrium, agent bias strengthens agent capacity investment incentives if and only if oversight also strengthens these incentives.*

Proof of Proposition 2. This follows from the fact that neither agent bias t_A nor the agent's aversion to being overturned π appear in equation 4, but both t_A and π appear in the agent's capacity investment given by equation 5. ■

Principal decision-making

Lemma 5. *When the agent will always be overturned by the overseer if he is authorized by the principal, the principal authorizes the agent to make policy if $c = 0$.*

Proof of Lemma 5. This follows from incentive compatibility for the principal to authorize the agent when he will always be overturned by the overseer. First, the principal's subjective expected payoff for not authorizing the agent is simply,

$$\begin{aligned} U_P(a=0; r=1, e^*) &= -y^2 - ca, \\ &= -\mathbb{E}[\omega]^2 - V[\omega] - c(0), \\ &= -V_F. \end{aligned}$$

Now, since if the principal authorizes the agent he will get overturned the policy payoff is the same but she must incur c. So, her subjective expected payoff from authorizing an agent in this environment is given by,

$$U_P(a=1; r=1, e^*) = -V_F - c.$$

Combining these two expected payoffs yields the principal's net expected payoff for authorizing the agent when the overseer is perfectly skeptical and is given by $\Delta U_P(a=1; r=1) = U_P(a=1; r=1) - U_P(a=0)$:

$$\begin{aligned} \Delta U_P(a=1; r=1) &= -V_F - c + V_F, \\ &= -c. \end{aligned}$$

Incentive compatibility implies that the principal, given her net expected payoff for doing so, will only authorize the agent to make policy when the overseer will overturn with certainty if $\Delta U_P(a = 1; r = 1) \geq 0$, which requires that $-c \geq 0$, or, equivalently, $c \leq 0$. Since $c \geq 0$ by assumption, the principal only delegates if $c = 0$, as stated in the result. ∎

Lemma 6. *When the agent will be upheld with certainty if authorized to make policy the principal authorizes the agent to make policy if* $t_A^2 \leq V_F - V_\varepsilon(\hat{e}_A(\beta, \kappa)) - c$.

Proof of Lemma 6. This follows from incentive compatibility for the principal to authorize the agent when the agent will always be upheld. First, the principal's subjective expected payoff when she does not authorize the agent to make policy is again,

$$U_P(a = 0) = -V_F.$$

Similarly, given that the principal knows that if she authorizes the agent to make policy then $x^*(\omega) = \omega + t_A$ and $e^* = \hat{e}_A(\beta, \kappa)$, her subjective expected payoff for authorizing the agent to make policy in this environment is given by,

$$\begin{aligned} U_P(a = 1; r = 0, e^* = \hat{e}_A(\beta, \kappa)) &= -t_A^2 - \mathbb{E}[\varepsilon | e^*]^2 - V[\varepsilon | e^*] - c, \\ &= -t_A^2 - V_\varepsilon(\hat{e}_A(\beta, \kappa)) - c. \end{aligned}$$

Combining these expected payoffs yields the principal's net expected payoff, defined as $\Delta U_P(a = 1; r = 0, \hat{e}_A(\beta, \kappa)) = U_P(a = 1; r = 0, e^* = \hat{e}_A(\beta, \kappa)) - U_P(a = 0)$:

$$\Delta U_P(a = 1; r = 0, \hat{e}_A(\beta, \kappa)) = -t_A^2 - V_\varepsilon(\hat{e}_A(\beta, \kappa)) - c + V_F.$$

Incentive compatibility implies that the principal will authorize the agent if and only if $\Delta U_P(a =$

$1; r=0, \hat{e}_A(\beta,\kappa)) \geq 0$, which requires that,

$$\begin{aligned} -t_A^2 - V_\varepsilon(\hat{e}_A(\beta,\kappa)) - c + V_F &\geq 0, \\ V_F - V_\varepsilon(\hat{e}_A(\beta,\kappa)) - c &\geq t_A^2, \end{aligned}$$

as stated in the lemma. ■

Proposition 3. *If the agent will either always be overturned or always be upheld following delegation, the principal never benefits from a biased agent relative to an ally agent.*

Proof of Proposition 3. First, we consider the case in which the agent, if authorized, will be overturned by the overseer with certainty. From lemma 5, the principal's net expected utility for authorizing the agent is given by,

$$\Delta U_P(a=1; r=1) = -c.$$

Since t_A does not appear in the principal's net expected utility, even if the principal benefits from empowering the agent (which *may* be the case when $c=0$) she does not benefit from a biased versus an ally agent.

Now we consider the second case: authorizing an agent that will receive perfect deference. Recall that if authorized, since the overseer will always uphold regardless of agent capacity investment, the agent makes capacity investment $\hat{e}_A(\beta,\kappa)$. From lemma 6, the principal's net expected payoff from authorizing the agent is given by,

$$\Delta U_P(a=1; r=0, \hat{e}_A(\beta,\kappa)) = -t_A^2 - V_\varepsilon(\hat{e}_A(\beta,\kappa)) - c + V_F.$$

Straightforward inspection of this net payoff reveals that the principal's utility is decreasing in agent bias $|t_A|$. As agent bias increases, the principal's utility decreases since no other component of her net expected utility is affected by changes in t_A. Thus, any time the principal benefits from delegating to an agent that will face perfect deference oversight, she strictly prefers an ally to biased agent. ■

Lemma 7. *Suppose the agent, if authorized, will make capacity investments sufficient to be upheld by the overseer. Further, suppose agent and overseer ideal points are organized such that the agent faces a conditional-deference overseer. Then the principal authorizes the agent to make policy if* $-2t_A t_R \geq c$.

Proof of Lemma 7. This follows from incentive compatibility for the principal to authorize the agent to make policy when facing a conditional-deference overseer. Note first that we are assuming that if the principal authorizes the agent to make policy then the agent will make capacity investments equal to the overseer's threshold $\underline{e}_R(t_A)$ and be upheld. The principal's subjective expected payoff if she does not authorize the agent is again,

$$U_P(a=0) = -V_F.$$

Now, given that the principal knows that if she authorizes the agent to make policy the agent will invest $e^* = \underline{e}_R(t_A)$, her subjective expected payoff for authorizing the agent is given by,

$$\begin{aligned} U_P(a=1;r=0,e^*=\underline{e}_R(t_A)) &= -(x^*(\omega)-\omega+\varepsilon)^2-c, \\ &= -t_A^2-\mathbb{E}[\varepsilon|e^*]-V[\varepsilon|e^*]-c, \\ &= -t_A^2-V_\varepsilon(\underline{e}_R(t_A))-c. \end{aligned}$$

Define the principal's net expected payoff from authorizing the agent as $\Delta U_P(a=1;r=0,\underline{e}_R(t_A)) = U_P(a=1;r=0,e^*=\underline{e}_R(t_A)) - U_P(a=0)$:

$$\Delta U_P(a=1;r=0,\underline{e}_R(t_A)) = -t_A^2-V_\varepsilon(\underline{e}_R(t_A))-c+V_F.$$

Incentive compatibility implies that the principal will authorize the agent to make policy if $\Delta U_P(a=$

$1; r = 0, \underline{e}_R(t_A)) \geq 0$, which requires that,

$$-t_A^2 - V_\varepsilon(\underline{e}_R(t_A)) - c + V_F \quad \geq \quad 0.$$

Now, solving the overseer's incentive compatibility condition to uphold with equality for $V_\varepsilon(e)$ allows us to substitute $V_\varepsilon(\underline{e}_R(t_A))$ as follows:

$$\begin{aligned} -t_A^2 - \left[V_F - t_A^2 + 2t_A t_R\right] - c + V_F &\geq 0, \\ -2t_A t_R &\geq c. \end{aligned}$$

This implies the principal will delegate if and only if $-2t_A t_R \geq c$, as stated in the result. ■

Proposition 4. *If the agent will make capacity investments at the threshold level required by the overseer following delegation, and authorization costs are positive, then the principal benefits from a biased agent relative to an ally whenever she benefits from delegating policymaking authority.*

Proof of Proposition 4. From lemma 7, if the principal benefits from authorizing the agent to make policy given that he will face conditional-deference oversight then,

$$-2t_A t_R \geq c.$$

First note that since $t_R < 0$ by assumption the principal's utility is increasing in $t_A > 0$. For the inequality to hold it must be that either $t_A > 0$ if $c > 0$ or $t_A \geq 0$ if $c = 0$. In either case, the principal is better off with a positively biased agent, $t_A > 0$, since her utility is increasing in t_A. Thus, any time the principal benefits from delegating to the agent (i.e., the above inequality holds) the principal also benefits from that agent being positively biased. ■