CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The University of Nottingham

Discussion Paper No. 2011-04

Vincent Anesi
and
Daniel Seidmann
April 2011

Optimal Delegation with a Finite Number of States

CeDEx Discussion Paper Series
ISSN 1749 - 3293

Sue Berry
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0) 115 95 15469
Fax: +44 (0) 115 95 14159
sue.berry@nottingham.ac.uk

Optimal Delegation with a Finite Number of States*

Vincent Anesi†
University of Nottingham

Daniel J. Seidmann‡
University of Nottingham

13th June 2011

Abstract

This paper studies delegation without monetary transfers when the number of possible states is small, and therefore finite. To do so, we fully characterize the class of optimal delegation sets in the finite-state version of Holmstrom's (1984) seminal model and analyze their properties. Our finite state assumption entails the following results: (i) the agent never takes her ideal decision, and takes a decision strictly between her and the principal's ideal (thus compromising with the latter) in low enough states; (ii) the agent takes the same decision in high enough states, and is indifferent between the decision she takes and the next highest decision in every other state; (iii) the agent may be induced to take decisions outside the support of the principal's ideal decisions; (iv) marginal increases in the agent's bias do not (generically) cause optimal delegation sets to shrink, and may increase the variance of the decision taken by the agent. We also show that the principal and the agent may both be better off if the latter cannot distinguish between some states.

Keywords: Optimal delegation, finite states, Ally Principle, expertise

1 Introduction

This paper studies delegation without money transfers when the number of possible states of the world is finite. The following example illustrates our model. It is widely believed that juries are unduly biased towards reaching some verdict (rather than hanging), and that they may therefore compromise by convicting on some lesser charge when evidence on

*We thank Claudio Mezzetti and Eyal Winter for their comments. We also appreciate the suggestions and comments of participants in seminars in Brussels, Haifa, Jerusalem, Madrid, Nottingham, Toulouse, and Warwick.

†School of Economics, Room B18, The Sir Clive Granger Building, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom. Email: vincent.anesi@nottingham.ac.uk.

‡School of Economics, Room B34, The Sir Clive Granger Building, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom. Email: daniel.seidmann@nottingham.ac.uk.

a more serious charge is unclear.[1] Accordingly, juries are instructed to convict a defendant on a charge only if they agree that the offense was committed.[2] For instance, if jurors believe that the defendant may have committed murder but definitely did not commit manslaughter then they may not convict on the lesser charge. The judge's instruction addresses the jury's bias, relative to the presumption of innocence, by restricting the verdicts which the jury may reach, inducing the jury to pool on acquittal unless the evidence proves guilt on some charge beyond a reasonable doubt.

One can think of this situation as an optimal delegation problem without monetary transfers in which the principal is the judge and the jury is the expert agent: expert, in the sense that its role is fact-finding. The uninformed principal (he) offers a delegation set – namely, a collection of scalar decisions (the permissible verdicts) – to the biased agent (she), who takes one of these decisions after observing the state (the evidence). All of these features appear in the literature on optimal delegation initiated by Holmstrom (1984). Almost all of this literature also supposes that the set of possible states is an interval and, therefore, an infinite set — we discuss the exceptions below. In the jury example, there may be many feasible decisions (lesser charges), but there is only a finite number of possible states: the evidence either clearly exonerates, or clearly inculpates the defendant of murder, or identifies the crime as murder but leaves the defendant's factual guilt in doubt.

The aim of this paper is to study optimal delegation when the number of states is small, and therefore finite. One might conjecture that the distinction between finite and an interval support of states is only a mathematical detail, and that no further insight can be gained by studying delegation models with a finite set of states. Our results reveal that such an impression is wrong. Analysis of the finite-state case is important not only because it describes situations that are clearly realistic, but also because finiteness of the set of states generates new and interesting properties of the optimal delegation set (ODS, for short). These turn on the principal's ability to exploit the distance between the agent's ideal decision across states.

In order to emphasize the role of a small number of states in optimal delegation, we analyze Holmstrom's (1984) seminal model of optimal delegation in an environment with a finite number of states. Holmstrom assumes that the principal's and the agent's preferences are each represented by loss functions which are quadratic around the player's ideal decision. The principal's ideal decision is simply the realization of the state, while the agent's ideal decision is b higher: $b > 0$ is the agent's *bias*. The agent observes the actual realization of the state; the principal does not, but believes that it is uniformly distributed on $[0, 1]$.[3] Holmstrom shows that the ODS then has a simple structure, which has been exploited in a number of applications. We extend Holmstrom's model by retaining all of

[1] See Hannaford-Agor et al (2002) pp 42-43 for field evidence in support of this possibility.

[2] See *Stein v. NY* 346 US (1953) on jury instructions and compromise verdicts. According to *Beck v. Alabama* 447 US (1980), juries need only be instructed to consider a lesser offense if there is evidence that it might have been committed. Hoffheimer (2006) discusses these issues.

[3] This is the benchmark example in Crawford and Sobel's (1982) cheap talk model. The optimal delegation problem is equivalent to a cheap talk game in which the Receiver moves first by committing to the decisions he will take in response to any message.

these assumptions except for the uniformly distributed state. Instead, we assume that the state is equally likely to take each of a finite number of equi-distanced values; so our model approximates Holmstrom's when we let the number of states tend to infinity. Their simplicity notwithstanding, maintaining Holmstrom's assumptions on preferences and the distribution of states has three major advantages. First, our approach allows us to derive a full and tractably simple characterization of the class of ODSs which, like Holmstrom's model, can be used in applications (cf. Sections 5 and 6 below). Second, and more importantly, some of the properties of the ODS we want to emphasize can be obtained by relaxing assumptions other than the interval support, but are never satisfied under Holmstrom's original model. Hence, retaining all of Holmstrom's assumptions except for the interval support enables us to identify finiteness of the number of states as the origin of our results. Finally, analysis of finite-state delegation problems naturally raises combinatorial issues. These issues can be surmounted with Holmstrom's assumptions: we not only fully characterize the ODS, but also show that it has a simple structure (which we describe below).[4]

Characterizing ODSs in this finite environment, we show that they satisfy the following properties:

Compromise. In every state, the agent takes a decision which is strictly less than her ideal decision in that state. In low enough states, the agent's decision exceeds the principal's decision in that state, and therefore represents a compromise between the two players' interests in such states. This stands in radical contrast to Holmstrom's (1984) result, where the agent takes her ideal decision in every low enough state.[5]

The chain property and top loading. In common with Holmstrom's model, the agent takes the same decision in all high enough states, and otherwise separates: a property which we call *top loading*. If the agent is biased enough then the ODS consists of a single decision; otherwise, and in contrast to Holmstrom (1984), the agent is indifferent in low states between the decision she takes and the next highest decision in the ODS (the *chain property*). The chain property implies that the principal must exclude some intermediate decisions (as in our example of judicial instructions) as well as some decisions above the ODS. Gailmard (2009) p26 notes that the FTC, like other agencies, is typically limited to up-down choices; Szalay (2005) p1174 mentions several other examples. By contrast, Holmstrom's principal need only exclude decisions above the ODS (cf. Melumad and Shibano (1991)).[6]

Extreme decisions. In Holmstrom's model, the principal only allows the agent to take decisions which are ideal for the principal in some state. (See Holmstrom (1984) and Melumad and Shibano (1991).). By contrast, we show that, for an interval of biases,

[4] The combinatorial issues resolve into a taxonomy which turns on the parities of the number of states and the number of decisions in the ODS. Analysis of delegation problems with interval support raises other issues, such as existence of an ODS (which is not an issue here). Such problems are also surmountable in Holmstrom's model.

[5] In both models, the agent may take a decision below the principal's ideal in high enough states: e.g. when the bias is so great that the ODS is a singleton.

[6] Martimort and Semenov (2006), Alonso and Matouschek (2008) and Gailmard (2009) provide other general conditions for this result; Kovac and Mylovanov (2009) and Goltsman et al (2009) prove that principal does not gain from offering mixtures of decisions in Holmstrom's benchmark model.

the agent takes a decision in the highest state which exceeds the principal's ideal in that (or any other) state. This result suggests a possible explanation for the tenure system: the university/principal forces the dean/agent to offer unduly generous terms to good candidates to motivate her to dismiss bad candidates.

Variations in bias and discretion. The theory also provides simple predictions about how variations in the bias affect the ODS. Unsurprisingly, a higher bias makes delegations less valuable to the principal. However, in contrast to Homstrom's model, increases in bias do not cause the ODS to shrink (except for non-generic values of the bias). Instead, a marginal increase in bias raises all decisions in the ODS whenever bias is low enough for delegation to be valuable, and high enough that the principal cannot achieve his first best. In addition, and again contrary to Holmstrom's model, the variance of decisions may increase with the bias.

Expertise. We end the paper by applying our results to a model in which the principal chooses the agent's expertise (or which agent to appoint) before offering a delegation set. A recent literature, which explores the trade-off between loyalty and competence, relies on the supposition that the principal prefers to appoint the more expert of two agents with a common bias (cf. Gailmard and Patty (2007) and Huber and McCarty (2004) Proposition 3). This property holds trivially in models (like Bendor and Meirowitz (2004)), which treat an amateur as an agent who does not observe the state with positive probability. By contrast, we treat an amateur as an agent who is unable to distinguish between a subset of states. The principal cannot gain by appointing an amateur if the bias is low enough because the expert would take almost first-best decisions in every state, whereas an amateur must be less state-sensitive. He can also not gain from appointing an amateur if the bias is large enough that delegation to an expert has no value. However, the property fails for every intermediate bias, in the sense that the principal strictly prefers to appoint some amateur over an expert with the same bias. This result holds whenever there are at least four states and, therefore, also holds in Holmstrom's model. Consequently, evidence that political appointees are less expert (cf. Lewis (2007)) does not imply that there is a trade-off between loyalty and competence.

If the expert can be outperformed, which amateur would the principal appoint from a pool of candidates with common bias? We address this question when there are up to five states. In all cases, the best amateur cannot distinguish between two succeeding states. With five states, this pair moves upward (towards higher states) and then downward as the bias increases.

We also consider an analogous question from the agent's point of view. Suppose that the agent can (publicly) commit not to distinguish between some states before being offered a delegation set. Consider, for example, a defense attorney (the agent) who observes the state by asking questions of her client, but has divergent interests (akin to the bias) in the sense that she must report intended perjury to the court. The attorney can then commit to being an amateur in our sense by not asking unduly detailed questions. We show that the principal and the agent may both prefer that the agent is some amateur rather than the expert for a range of bias, and that the same amateur may be best for both the principal and the agent. On the other hand, computations for low numbers of states reveal that the agent prefers to be an expert when the principal would appoint an expert, but not conversely.

Related literature

Although the literature on optimal delegation is now too large to survey exhaustively, we briefly explain the paper's relationship to a few of the most closely related game-theoretic contributions.

We have already mentioned the main results from Holmstrom's (1984) interval support model as benchmarks against which to gauge the impact of a finite state variable in our model. Melumad and Shibano (1991) generalize Holmstrom's results proving, inter alia, that delegation is an optimal mechanism: a property which also holds in our model.

Various other papers have characterized the ODS in variants on Holmstrom's model (with interval support), demonstrating that some of his conclusions may fail under different assumptions. Both Alonso and Matouschek (2008) and Szalay (2005) show that the principal may exclude some intermediate decisions when the support of states is an interval. Alonso and Matouschek (2008) show that the ODS might not be an interval with more general preferences if the agent's ideal decisions are insufficiently state-sensitive. This motive is absent in our model because a state-independent bias precludes any conflict over state-sensitivity. In Szalay (2005), the principal may exclude intermediate decisions to induce costly information acquisition, whereas expertise is fixed in our main model. Decisions are two-dimensional in Koessler and Martimort (2009), but the support of states is an interval. Unlike the principal, the agent has different biases in each dimension, which allows the principal to screen the agent by distorting each dimension away from the agent's ideal decision. They show that an agent with any bias takes a different decision-pair in each state, but does not take her ideal decision on either dimension in any state. By contrast, our model (and Holmstrom's) precludes such screening. The agent does not take her ideal decisions because the ODS induces her to compromise with the principal in every state when the support is finite.

The only previous models with a finite number of states are Green (1982), Huber and Gordon (2007) and Amador et al (2006). Green shows that the optimal stochastic delegation set is the solution to a linear programming problem if (contrary to our model) there is a finite number of feasible decisions. Huber and Gordon characterize the ODS for an example with three feasible decisions. Amador et al show that the principal might achieve first best in a two-state model if preferences are sufficiently aligned and any scalar decision is feasible (as in our model); that the ODS features separation with intermediate alignment; and that delegation is otherwise not valuable. We generalize these properties to any finite number of states.

Other notable, but less related contributions include Ambrus and Egorov (2009), Armstrong and Vickers (2010), Krishna and Morgan (2008) and Mylovanov (2008).

Our discussion of the principal's choice of an agent is closely related to Ivanov (2010), who constructs an amateur who outperforms the expert in Holmstrom's model whenever delegation is valuable.[7] We allow for a richer set of possible amateurs, which we exploit

[7] The Expertise Principle also fails in Postlewaite's (1982) Case II; but preferences therein fail the single crossing property, and there is a finite number of feasible decisions.

when characterizing the best agent. Our demonstration that the principal may prefer to appoint an amateur is reminiscent of a result in the signalling literature: the principal may prefer to take advice from an amateur: cf. Fischer and Stocken (2001) and Ivanov (2010).

The agent's choice of expertise is related to the problem studied by Szalay (2005) inter alia, where the agent can observe the state exactly at some private cost of effort. By contrast, we suppose that acquiring expertise is costless, and that the principal offers a delegation set after observing the agent's expertise.

In Section 2, we define the basic model used to describe optimal delegation problem. In Section 3 we analyze that problem and derive the ODS. We discuss the general properties of the ODS in Sections 4 and 5. The application to expertise is presented Section 6. We summarize and discuss how our results might generalize in Section 7, and provide lengthier proofs in the Appendix and an online Appendix (Anesi and Seidmann (2011)).

2 The model

Players and preferences

There are two players in the model: a principal (he) and an agent (she). The preferences of both players depend on a decision variable, d, and on a random state of the world, t, in which the decision is taken.[8]. We assume that there is only a finite set of conceivable states, which we denote $\mathbf{T} \equiv \{0, 1, \ldots, T-1\}$, but that d can take any real value. The principal's evaluation of the decision d is represented by the following loss function:

$$\lambda(d,t) \equiv \left(d - \frac{t}{T-1}\right)^2 .$$

A principal who knew the state t for sure would therefore take decision $d = t/(T-1)$. Accordingly, we refer to $t/(T-1)$ as the principal's ideal decision in state t. We normalize by dividing the state by $T-1$ in order to facilitate comparison with Holmstrom's (1984) continuous-state model.

The agent's evaluation of decision d is given by loss function

$$\Lambda(d,t) \equiv \left(d - b - \frac{t}{T-1}\right)^2$$

for some $b > 0$. For any value of b, the agent's ideal decision in state t is thus $b + t/(T-1)$, which exceeds the principal's ideal decision in state t by b. We refer to b as the agent's *bias*. By construction, the bias is state-independent.

An obvious, but important, implication of these loss functions is that the principal strictly prefers the agent to take decision d over decision e in those states where the agent weakly prefers d over e. This property will play a decisive role in the analysis below.

[8]There are no money transfers, as in Holmstrom (1984). We discuss the implications at the end of Section 3.2

Information and timing

Events unfold as follows: 1) Nature chooses a state t in $\mathbf{T}$ and – unless otherwise stated – reveals the true value of t privately to the agent; 2) The principal offers the agent a delegation set (to be defined shortly); 3) The agent takes a decision d in the delegation set.

The agent is asymmetrically informed about the state relative to the principal, who only knows that every possible state is equally likely. Both players' loss functions are common knowledge.

Delegation sets

Broadly defined, a delegation set is any collection of decisions: that is, any nonempty subset of $\mathbb{R}$. A delegation set (say, Δ) could include decisions which the agent would never take. We follow the literature by focusing on minimal delegation sets: that is, delegation sets with the property that a loss-minimizing agent with bias b would take each decision in Δ in some state. This is the right focus because an optimal delegation set minimizes the principal's expected loss over all incentive compatible mechanisms (cf. Melumad and Shibano (1991)); so exactly the decisions in a minimal delegation set are taken according to every optimal mechanism. Accordingly, we define $\mathcal{D}(b)$ as the class of minimal delegation sets when the agent's bias is b. We will henceforth drop the qualifier 'minimal'.

A brief inspection of λ^A reveals that an agent who is offered $\Delta \in \mathcal{D}(b)$ can only be indifferent between two distinct decisions in any state; so a delegation set must contain a finite number of decisions. We will suppose that the agent takes the lowest decision in Δ which minimizes her loss in each state.[9] Formally, for every state $t \in \mathbf{T}$, define the decision function $d_t : \mathcal{D}(b) \times \mathbb{R}_+ \to \mathbb{R}$ as

$$d_t(\Delta, b) \equiv \min\left\{ \arg\min\left\{ \left(d - \frac{t}{T-1} - b\right)^2 : d \in \Delta \right\}\right\} .$$

Note that we implicitly assume here that the agent never mixes over decisions. This is without loss of generality because it is never optimal for the principal to induce the agent to randomize.

We say that $\Delta \in \mathcal{D}(b)$ is an *optimal delegation set* (or *ODS*) if and only if

$$\Delta \in \arg\min\left\{ \sum_{\tau \in \mathbf{T}} \left(d_\tau(\Phi, b) - \frac{\tau}{T-1}\right)^2 : \Phi \in \mathcal{D}(b) \right\} .$$

The class of optimal delegation sets when the agent's bias is b is denoted by $\mathcal{D}^*_T(b)$. A generic element of $\mathcal{D}^*_T(b)$ is of the form $\{\delta_0, \ldots, \delta_K\}$, where we order decisions such that $i < j$ implies $\delta_i < \delta_j$. In the next section, we characterize the class of ODSs for every bias.

Some more terminology will prove useful. A principal who offers a single-decision ODS must offer the decision which he would take if he did not delegate. Accordingly, we will say that *delegation is valuable* if and only if every ODS contains more than one decision.

[9] Single crossing implies that this supposition is without loss of generality.

It will also be useful to benchmark our results against an alternative model in which the principal places no restrictions on the agent (as in Dessein (2002)), implicitly offering the delegation set consisting of decisions $\left\{b + \frac{t}{T-1}\right\}$ for $t \in \mathbf{T}$. We will say that such a principal *fully delegates*.

3 Derivation of the optimal delegation set

3.1 Top loading and the chain property

It is important for both theoretical and practical purposes that ODSs have a simple structure. It turns out that every ODS satisfies two simple conditions. We need some more notation to specify these conditions. For every $\Delta = \{\delta_0, \dots, \delta_K\} \in \mathcal{D}(b)$, define $t_k(\Delta, b)$ and $T_k(\Delta, b)$ as follows:

$$T_k(\Delta, b) \equiv \{\tau \in \mathbf{T} : d_\tau(\Delta, b) = \delta_k\} \ ,$$
$$t_k(\Delta, b) \equiv \max\{\tau : \tau \in T_k(\Delta, b)\} \ .$$

The elements of $T_k(\Delta, b)$ are the states in which the agent takes decision δ_k from Δ, and $t_k(\Delta, b)$ is the maximal such state.

Definition 1. *A delegation set $\Delta = \{\delta_0, \dots, \delta_K\} \in \mathcal{D}(b)$ satisfies top loading if and only if $|T_k(\Delta, b)| > 1$ implies that $k = K$.*

Definition 2. *A delegation set $\Delta \in \mathcal{D}(b)$ satisfies the chain property if and only if $\tau_1 = t_k(\Delta, b)$ and $\tau_2 = t_{k+1}(\Delta, b)$ imply that*

$$\Lambda\left(d_{\tau_1}(\Delta, b), \tau_1\right) = \Lambda\left(d_{\tau_2}(\Delta, b), \tau_1\right)$$

for every $k = 0, \dots, K-1$.

In words: Δ is top loaded when only the largest decision in Δ is taken in more than one state; and Δ satisfies the chain property if an agent who takes different decisions in states t and $t+1$ is indifferent between them in state t. If Δ satisfies the chain property then it must exclude any decisions between those taken in states t and $t+1$ because convexity of the agent's loss function implies that she would prefer this compromise in state t.[10] Our example of jury instructions in the Introduction illustrates a delegation set which excludes compromises.

Theorem 1 below asserts that every ODS satisfies the chain property and top loading. The ODS in Holmstrom's interval support model is top loaded, but does not satisfy the chain property.[11] Cheap talk models with interval support satisfy an analog of the chain property: the support is partitioned into a finite number of intervals in every equilibrium; and, at the maximal state in all but the largest element of the partition, the agent is

[10] This argument implies that a non-minimal delegation set which satisfied the chain property would also have to exclude compromises.

[11] The ODS fails top loading in Koessler and Martimort (2009).

indifferent between the decision it induces and the next highest decision taken by the principal.[12] However, these equilibria do not satisfy top loading.

An important implication of Definitions 1 and 2 is that any delegation set Δ which is top loaded and satisfies the chain property, is defined by two parameters: the lowest decision in Δ (namely, δ_0) and the number of decisions it contains (namely, $K+1$). Theorem 1 establishes that the number of decisions in an ODS for an agent with bias b is determined by the values of K which satisfy

$$b \in \left[b^{\min}(K,T), b^{\max}(K,T)\right]$$

where, for all integers $s \le T$ and $K \le s-1$:[13]

$$b^{\min}(K,s) \equiv \frac{s-K}{2(T-1)} \text{ and } b^{\max}(K,s) \equiv \begin{cases} \frac{s-K+1}{2(T-1)} & \text{if } K > 0 \ , \\ +\infty & \text{otherwise.} \end{cases}$$

This result immediately implies that: (i) there is a unique ODS if $b \neq \frac{k}{2(T-1)}$ for any $k \in \{1,...,T-1\}$, and only two ODSs otherwise; and (ii) delegation is not valuable if and only if $b > b^{\min}(0,T) = \frac{T}{2(T-1)}$.

Having determined K, the lowest decision in the ODS is then chosen to minimize the principal's expected loss, subject to the agent being indifferent between δ_t and δ_{t+1} in every state $t \in \{0,\ldots,K-1\}$, and taking the maximal decision in every higher state.

3.2 The ODS

We are now in a position to fully characterize the ODS:

Theorem 1. *If $b \le \frac{1}{2(T-1)}$ then the ODS consists of the T decisions which are ideal for the principal in each state: $\Delta^* = \left\{\frac{t}{T-1}\right\}_{t\in\mathbf{T}}$.*

If $b > \frac{1}{2(T-1)}$ then

(i) Every ODS satisfies the chain property and top loading;

(ii) If $b \neq \frac{k}{2(T-1)}$ for any $k \in \{2,...,T-1\}$ then there is a unique ODS, $\Delta^* \equiv \{\delta_0^*,\ldots,\delta_K^*\}$. The number of decisions in the ODS, $K+1$, is implicitly determined by $b \in \left(b^{\min}(K,T), b^{\max}(K,T)\right)$; and

$$\delta_0^* = \begin{cases} \frac{K}{T}b - \frac{1}{2(T-1)} + \frac{(T-K)^2}{2T(T-1)} & \textit{if } K \textit{ is even,} \\ \frac{2T-K-1}{T}b - \frac{T-1}{2T(T-1)} - \frac{(T-K)^2}{2T(T-1)} & \textit{if } K \textit{ is odd.} \end{cases}$$

In particular, if $b > \frac{T}{2(T-1)}$ then delegation is not valuable: $\Delta^* = \left\{\frac{1}{2}\right\}$.

(iii) If $b = \frac{k}{2(T-1)}$ for some $k \in \{2,...,T-1\}$ then there are two ODSs: the ODS with $T-k$ and with $T-k+1$ decisions defined in part (ii) above.

[12] This analog of the chain property does not hold in cheap talk games with a finite support.

[13] We introduce slightly more general notation for use in the next subsection.

Part (ii) implies that the i^{th} decision in the ODS is

$$\delta_i = \begin{cases} \ell(b,T,K) + \frac{i}{T-1} & \text{if } i \text{ is even,} \\ 2b - \ell(b,T,K) + \frac{i-1}{T-1} & \text{if } i \text{ is odd,} \end{cases}$$

where

$$\text{and } \ell(b,T,K) \equiv \begin{cases} \ell^e(b,T,K) & \text{if } K \text{ is even,} \\ \ell^o(b,T,K) & \text{if } K \text{ is odd.} \end{cases}$$

and

$$\ell^e(b,T,K) \equiv \frac{K}{T}b - \frac{1}{2(T-1)} + \frac{(T-K)^2}{2T(T-1)} \ ,$$
$$\ell^o(b,T,K) \equiv \frac{2T-K-1}{T}b - \frac{1}{2T} - \frac{(T-K)^2}{2T(T-1)} \ , \qquad (1)$$

We prove Theorem 1 in Appendix A. We will explain how we prove the result in the next subsection. Appendix C illustrates the ODS for the special cases of three, four and five states.

The principal's problem could be solved by finding the loss-minimizing delegation set with K decisions, and then minimizing again over K. While Theorem 1 asserts that every ODS satisfies top loading, this is not true of the best delegation sets containing a fixed number of decisions. For example, if there are four states and $\frac{1}{6} < b < \frac{1}{3}$ then the best two-decision delegation set is $\{b, b + \frac{2}{3}\}$, which fails top loading because the agent takes decision b in states 0 and 1. However, the ODS contains three decisions and satisfies top loading when $\frac{1}{6} < b < \frac{1}{3}$.

Our assumption that money transfers are unavailable does not seem to qualitatively affect our results: we can show that the ODS satisfies the chain property and top loading when there are three states and a delegation set consists of money-decision pairs.[14]

3.3 Proof strategy

The technical problem in determining the ODS in models with interval support is how to convert it into a finite-dimensional problem. On the other hand, Melumad and Shibano (1991) Lemma 1 demonstrates that incentive constraints are very restrictive in models with interval support: the agent takes her ideal decision if the decisions taken are locally continuous and strictly monotonic in the state. Alonso and Matouschek (2008) use this property to provide conditions for the principal to reduce his loss by adding or removing decisions from a given delegation set. In particular, adding an intermediate decision to any finite delegation set induces the agent to take a less extreme decision in some intermediate states. Alonso and Matouschek use such local patches (adding or removing intermediate decisions) to characterize the ODS.

Models with finite support are automatically finite-dimensional. However, the incentive constraints are no longer as strong because the agent's ideal decisions are $1/(T-1)$ apart; so

[14] Details are available from the authors on request.

the effect of adding or removing intermediate decisions depends on the original delegation set. In particular, adding a decision to those taken in states t and $t+1$ can induce the agent to take this compromise decision in states $s \leq t$ (if the delegation set satisfies the chain property) or in states $s \geq t+1$ (if the agent strictly prefers her decision in state t) or in no states. In sum, we cannot use Alonso and Matouschek's local patches; so we use entirely different arguments to characterize the ODS.

One might conjecture that Theorem 1 could be proved by induction on the number of states, but this turns out not to be fruitful. Furthermore, as we noted above, the best K-decision delegation sets do not necessarily satisfy the chain property or top loading; so induction on K is uninstructive.

Instead, we exploit the recursive structure of the ODS. Consider the event $E \equiv \{\tau \geq t\}$ for some state $t > 0$. In light of single-crossing, the ODS induces the agent to take decisions which minimize the principal's loss on E subject to the agent preferring d_τ over $d_{\tau+1}$ for every state in $E \cup \{t-1\}$. We exploit this property by defining and characterizing ODSs on connected subsets of states, and use induction arguments on the size of these events. We will further exploit our characterization when we describe the ODS for amateur agents in Section 5.

We need some additional notation for this purpose. Let $\mathbf{T}_{s,t} \equiv \{t, t+1, \ldots, t+s-1\} \subseteq \mathbf{T}$, and define the probability distribution $p_{s,t}$ as

$$p_{s,t}(\{\tau\}) \equiv \begin{cases} 1/s & \text{if } \tau \in \mathbf{T}_{s,t} \\ 0 & \text{otherwise,} \end{cases}$$

for every $\tau \in \mathbf{T}$, where $\mathbf{T}_{T,0} = \mathbf{T}$. Thus, assuming that the state is distributed according to $p_{s,t}$ amounts to assuming that only the states in $\mathbf{T}_{s,t}$ can occur, each with the same probability.

We now generalize the definitions of Section 2, using this interpretation. Let $\mathcal{D}_{s,t}(b)$ be the class of delegation sets when the agent's bias is b and the state is in $\mathbf{T}_{s,t}$, and say that $\Delta \in \mathcal{D}_{s,t}(b)$ is an *ODS for* $\mathbf{T}_{s,t}$ if and only if

$$\Delta \in \arg\min \left\{ \sum_{\tau \in \mathbf{T}_{s,t}} p_{s,t}(\{\tau\}) \left(d_\tau(\Phi, b) - \frac{\tau}{T-1} \right)^2 : \Phi \in \mathcal{D}_{s,t}(b) \right\} .$$

The class of ODSs for $\mathbf{T}_{s,t}$ when the agent's bias is b is denoted by $\mathcal{D}^*_{s,t}(b)$.

We are now in a position to provide an outline of the construction on which the proof of Theorem 1 is based. The following existence result proves necessary as a first step to doing so because we will use negative arguments to characterize ODSs.

Lemma 1. *$\mathcal{D}^*_{s,t}(b) \neq \emptyset$ for every $t \in \mathbf{T}$, every integer $s \leq T - t$, and any $b \geq 0$.*

Lemma 1 establishes existence. The characterization argument is rather long, and involves three intermediate steps. These steps, however, are of some interest in their own right, as they expose interesting relationships between the various properties of ODSs.

Our next result relies on the following notation:

$$\ell^e(b,s,K) \equiv \frac{K}{s}b - \frac{1}{2(T-1)} + \frac{(s-K)^2}{2s(T-1)} \ , \tag{2}$$

$$\ell^o(b,s,K) \equiv \frac{2s-K-1}{s}b - \frac{s-1}{2s(T-1)} - \frac{(s-K)^2}{2s(T-1)} \ , \tag{3}$$

$$\text{and } \ell(b,s,K) \equiv \begin{cases} \ell^e(b,s,K) & \text{if } K \text{ is even,} \\ \ell^o(b,s,K) & \text{if } K \text{ is odd.} \end{cases} \tag{4}$$

Lemma 2. *Let $b > \frac{1}{2(T-1)}$. If $\Delta \equiv \{\delta_t, \ldots, \delta_{t+K}\}$ is in $\mathcal{D}^*_{s,t}(b)$ and satisfies the chain property and top loading then*

$$\delta_{t+\tau} = \begin{cases} \ell(b,s,K) + \frac{t+\tau}{T-1} & \textit{if } \tau \textit{ is even,} \\ 2b - \ell(b,s,K) + \frac{t+\tau-1}{T-1} & \textit{if } \tau \textit{ is odd,} \end{cases}$$

and K is implicitly determined by $b \in \left[b^{\min}(K,s), b^{\max}(K,s)\right]$.

Note that substituting $s = T$ into Lemma 2 yields the ODS identified in Theorem 1. The next step proceeds by induction on the number of states. We start from the following hypothesis:

($\mathbf{H_s}$) For every $z \leq s$ and every $t \leq T - z$, every $\Delta \in \mathcal{D}^*_{z,t}(b)$ satisfies the chain property and top loading.

($\mathbf{H_s}$) says that ODSs satisfy the chain property and top loading when the number of possible states does not exceed s. It is easy to confirm that, for every $t \in \mathbf{T} \setminus \{T-1\}$, an ODS for $\mathbf{T}_{2,t}$ satisfies the chain property and top loading: that is, ($\mathbf{H_2}$) is true.

($\mathbf{H_s}$) implies that every ODS for $s+1$ states has also the chain property:

Lemma 3. *If ($\mathbf{H_s}$) then, for every $t \leq T - s - 1$, $\Delta \in \mathcal{D}^*_{s+1,t}(b)$ implies that Δ satisfies the chain property.*

The last step asserts that the chain property and optimality jointly imply top loading. Formally:

Lemma 4. *If $\Delta \in \mathcal{D}^*_{s,t}(b)$ satisfies the chain property then Δ is top loaded.*

Lemmas 2, 3, and 4 jointly allow us to provide the full characterization of ODSs in Theorem 1.

4 Analysis of the ODS

Properties of the ODS in finite environments

We can now use Theorem 1 to study how a limited number of states affects the structure of optimal delegation sets. A useful benchmark against which to measure the impact of finite states is the continuous state model. Let $\mathcal{D}^*_\infty(b)$ be the class of ODSs when the agent's bias is b and the state is uniformly distributed on $[0,1]$ (viz. Holmstrom's model). Melumad and Shibano (1991) prove that $\mathcal{D}^*_\infty(b)$ consists of the unique delegation set

$$\Delta^*_\infty(b) \equiv \begin{cases} [b, 1-b] & \text{if } b \in [0, 1/2] \ , \\ \{1/2\} & \text{if } b > 1/2 \ . \end{cases} \tag{5}$$

The next result demonstrates that finiteness generates interesting features of optimal delegation that Holmstrom's model does not capture:

Corollary 1. *If Δ is an ODS, then the following statements are true:*
(i) $d_t(\Delta, b) > \frac{t}{T-1}$ for every $t \leq |\Delta| - 1$ if and only if $b > \frac{1}{2(T-1)}$;
(ii) $d_t(\Delta, b) < b + \frac{t}{T-1}$ for every $t \in T$;
(iii) Delegation is valuable if and only if $b < \frac{T}{2(T-1)}$;
(iv) At critical levels of the bias, $b = \frac{k}{2(T-1)}$, the two ODSs induce the agent to take the same decision in each state $t \leq T - k - 1$;
(v) The agent takes a decision below b if and only if delegation is valuable and the state is 0;
(vi) The agent takes a decision above 1 if and only if $\frac{1}{2(T-1)} < b < \frac{1}{T-1}$ and the state is $T-1$.

Part (i) asserts that the agent takes decisions which exceed the principal's ideal in all states where the chain property holds. If $b \leq 1/2(T-1)$ then the chain property does not apply because the ODS is first best for the principal. For larger biases, the ODS induces the agent to take a decision above the principal's ideal in states $t \leq K$. However, part (vi) implies that this property does not generalize to all states: the agent takes a decision less than 1 in state $T-1$ whenever $b > 1/(T-1)$.

We prove Corollary 1 in Appendix B. Three notable properties of ODSs emerge from our analysis:

(a) Compromise: Parts (i) and (ii) imply that the agent takes a decision strictly between her and the principal's ideal whenever $t \leq K$ and the ODS is not first best. In contrast to full delegation problems (cf. Dessein (2002)), the agent takes a decision strictly below her ideal in *every* state.

(b) Exclusion of intermediate decisions: Part (ii) also implies that intermediate decisions must be excluded from the ODS, even if we drop minimality: for every decision δ in the ODS Δ, there is a neighborhood of δ, $N(\delta)$, such that adding any element of $N(\delta)$ to Δ would increase the principal's loss.

(c) Extreme decisions: In full delegation problems, the agent necessarily takes her ideal decision in every state, irrespective of the cardinality of the support; so the minimal and maximal decisions are respectively b and $1+b$. In particular the agent takes decisions which exceed 1 (the maximal ideal decision for the principal) in every state $t > 1 - b$.

Part (v) asserts that the agent only takes a decision below b when the ODS is non-singleton. Theorem 1 then implies that the agent is indifferent in state 0 between the lowest two decisions in the ODS, which implies that she must take decisions above b in every positive state.

If the ODS contains T decisions but is not first best then parts (i) and (ii) imply that the agent takes a decision exceeding 1 in state $T-1$, just as in full delegation problems. However, in contrast to such problems, part (vi) asserts that the agent only takes a decision exceeding 1 in these circumstances. The intuition is that the agent is indifferent in state $T-2$ between the two highest decisions in an ODS which contains T decisions. This is only possible when the second highest decision is less than 1 because $b < 1/(T-1)$. Part (iv) implies that the ODS drops the highest decision when the bias increases above $1/(T-1)$, which implies part (vi).

Comparison with the interval support model

A brief inspection of $\Delta^*_\infty(b)$ reveals that none of these properties hold in the continuous-state version of the model. The agent never takes a decision below her lowest ideal decision (i.e., b) or above the principal's highest ideal decision (i.e., 1). Furthermore, if $b < 1/2$ then the agent takes her ideal decision in every state below $1-2b$: the ODS is an interval, so the principal does not exclude intermediate decisions.

Note that properties (a)-(c) can also be obtained by relaxing assumptions *other* than continuous states. For instance, the agent never takes her ideal decision in Koessler and Martimort (2009) because decisions are multi-dimensional. Considering more general payoff functions and state distributions, Alonso and Matouschek (2008) derive necessary and sufficient conditions for the ODS to include decisions inside the agent's range of ideal decisions. They also show that the ODS contains at most one decision below (resp. above) the agent's (resp principal's) range of ideal policies. As the comparison between $\mathcal{D}^*_T(b)$ ($T < \infty$) and $\mathcal{D}^*_\infty(b)$ makes apparent, however, having a small number of states suffices to generate (a)-(c).

If T is sufficiently large then the support of states is close to $[0,1]$, and the probability distribution is almost uniform. Although our focus is on the small-T case, it is of interest to note that the ODS in our model approaches the ODS in the continuous-state model as T becomes arbitrarily large:

Observation 1. *For all $b \geq 0$, $\lim_{T\to\infty} \mathcal{D}^*_T(b) = \mathcal{D}^*_\infty(b)$.*

We prove Observation 1 in Appendix B.

5 Variations in bias

We now conduct a series of comparative statics exercises on the ODSs derived in Section 3. Specifically, we investigate how changes in bias affect the distribution of decisions taken by the agent and, thereby, the principal's expected loss.

Optimal delegation and the agent's discretion

We first investigate how the ODS changes with the bias. For expositional convenience, define

$$\Delta : \mathbb{R}_+ \setminus \left\{ \frac{T-K}{2(T-1)} \right\}_{K=0}^{K=T-1} \to \mathcal{D}_T^*(b)$$

as the mapping that assigns the unique ODS to every non-critical value of the bias. Furthermore, describe any $b = \frac{k}{2(T-1)}$ for some $k \in \{1, ..., T\}$ as 'critical'. Comparative static analysis of $\Delta(b)$ yields:

Proposition 1. *Suppose that the initial bias is low enough that delegation is valuable.*

(i) Consider a marginal increase in bias from b_0 *to* b_1*. If* b_0 *is not critical then the principal raises all decisions in the ODS. By contrast, if* b_0 *is critical then the principal drops the highest decision in the ODS: viz.* $\lim_{b \searrow b_0} \Delta(b) \subset \lim_{b \nearrow b_0} \Delta(b)$.

(ii) Consider an increase in bias from b_0 *to* b_1 *which is large enough that* $\Delta(b_1)$ *contains fewer decisions than* $\Delta(b_0)$*. Then,*

$$d_0(\Delta(b_0), b_0) < d_0(\Delta(b_1), b_1) \ , \text{ and } d_{T-1}(\Delta(b_1), b_1) < d_{T-1}(\Delta(b_0), b_0)$$

(so $\Delta(b_0)$ *has a larger diameter than* $\Delta(b_1)$*).*

Proposition 1(i) implies that a small enough increase in b from a non-critical starting point shifts the ODS to the right, just as in models of full delegation (Dessein (2002)). In our model, the two players compromise on the decision taken in each state. A small increase in bias then raises each compromise decision.

Relatedly, Proposition 1 implies that $\Delta(b_0)$ and $\Delta(b_1)$ are only ordered by set-inclusion at critical values of the bias. This observation bears on a standard prediction in the literature on legislative control agencies, the Ally Principle, which asserts that the principal gives more discretion to a less biased agent. The literature defines discretion in terms of set inclusion: the principal gives less discretion to agent 0 than to agent 1 if he offers Δ_i to agent i and $\Delta_0 \subset \Delta_1$. Although set-inclusion is only a partial ordering of delegation sets, a brief inspection of (5) reveals that the Ally Principle holds in Holmstrom's (1984) continuous-state model.

Proposition 1 implies that this conclusion does not carry over to the finite state model when the bias is not critical.[15] This is not surprising, as set inclusion seems to be quite a demanding criterion in a finite-state environment. It is therefore striking that the Ally Principle holds in our model for marginal variations in bias at critical levels.

Discretion can alternatively be measured by the variance of the decision taken by the agent. The Ally Principle would then assert that the variance of the ODS is greater, the less biased is the agent. This version of the Ally Principle is evidently less demanding in a finite-state world.

[15] The previous literature (e.g., Huber and McCarty (2004) and Alonso and Matouschek (2008)) has already shown that relaxing some of Holmstrom's assumptions, other than interval support, may cause the Ally Principle to fail.

Formally, for every non-critical value of b, let

$$\mathbb{V}(b) \equiv \frac{1}{T}\sum_{t=0}^{T-1}\left[d_t\left(\Delta(b), b\right) - \frac{1}{T}\sum_{\tau=0}^{T-1} d_\tau\left(\Delta(b), b\right)\right]^2 .$$

We can use our characterization of the ODS to prove that

Observation 2. $\mathbb{V}$ *is strictly increasing (resp. constant) on* $\left(\frac{1}{2(T-1)}, \frac{1}{T-1}\right)$ *when* T *is odd (resp. even).*[16]

Observation 2 implies that this version of the Ally Principle fails for an interval of bias. Observation 2 may also hold for larger bias: the variance of decisions is strictly increasing in b when $K = 2$ and when $K = T - 3$ (but is decreasing when $K = T - 2$).

Observation 2 is a direct implication of finiteness: it is readily checked that the variance of the decision in Holmstrom's continuous-state model is strictly *decreasing* in b whenever delegation is valuable.[17]

Observation 2 describes the local behavior of the variance at non-critical bias. The variance of decisions decreases when b crosses a critical threshold.

The value of delegation

We now turn to another version of the Ally Principle in the literature: that the principal prefers to appoint an agent with a lower bias. The literature suggests various reasons why this might fail: for example, the principal may optimally appoint a more biased agent if her decision is itself a move in a larger game, as in Vickers (1985), where other players best-respond to the agent; or to motivate an agent to become better informed, as in Callander (2008). We now use Corollary 1 to show that this version of the Ally Principle holds in our model:

Proposition 2. *The principal strictly prefers to appoint a less biased agent.*

Proof: Let $\Delta_1 = \{\delta_0, \ldots, \delta_K\} \in \mathcal{D}_T^*(b_1)$. Theorem 1 implies that an agent with bias b_1 is indifferent between δ_t and δ_{t+1} in states $t < K$. An agent with bias $b_0 < b_1$ must strictly prefer δ_t over δ_{t+1} in states $t < K$; so the principal loses no more by offering Δ_1 to agent b_0 than he loses when offering (the optimal) Δ_1 to agent b_1.

Corollary 1 implies that $\delta_0 = d_0\left(\Delta_1, b\right) > 0$. The principal can then improve upon Δ_1 when transacting with the agent b_0 by reducing δ_0 towards 0.

□

Put differently, the value of delegation — i.e. the difference between the principal's (minimal) expected loss when he takes the decision himself and his expected loss when he delegates — decreases with the agent's bias.

[16] See Anesi and Seidmann (2011) for the requisite calculations.

[17] In further contrast, the expected decision is an inverted U-shaped function of bias in Holmstrom (1984); whereas a local increase in bias generically raises the entire ODS here (cf. Proposition 1).

6 Application: Expertise

Thus far, we have studied the constraints that the presence of preference divergence and asymmetric information between the principal and agent puts on the ODS. Our analysis up to this point has assumed that the agent is an expert in the sense that she perfect observes the realization of the state. In many real-world situations, however, principals seem not to appoint expert agents. According to many commentators, political appointments to bureaucracies have become increasingly prevalent in the US and the UK. Hurricane Katrina focused concerns that these bureaucrats are appointed for their loyalty, even though they are (on average) less competent than career civil servants.[18] A recent literature has built on these features by studying the ensuing trade-off between loyalty and expertise: cf. Egorov and Sonin (forthcoming), Gailmard and Patty (2007) and Huber and McCarty (2004).

In this section, we use our results from the previous sections to investigate whether the principal appoints the most expert agent when given the choice among a pool of agents with different levels of knowledge about the state but the same bias, and whether an agent would choose ex ante to become fully informed before she is offered a delegation set.

Like Ivanov (2010), we treat knowledge as partitional. Specifically, we suppose that each agent has a partition, $\{P_1, ..., P_n\}$, of the T states into $n \leq T$ events. The agent in previous sections has $n = T$; we refer to such an agent as the *expert*. In contrast to Ivanov, we do not require that the events are connected, in the sense that an agent may be unable to distinguish between states $t-1$ and $t+1$, but can distinguish state t. (States are ordered by their payoff implications.)[19]

We will apply our techniques and characterization results from the previous sections to study the payoff consequences of the agent having a coarser partition of the set of states (i.e. $n < T$). We explore the conditions on bias under which the principal and/or the agent can gain if the principal appoints an amateur rather than an expert. At first sight, it seems implausible that an amateur could outperform the expert because an amateur who cannot distinguish between states in some event must take the same decision in that event. The decision taken by an expert must therefore be at least as state-dependent as that taken by an amateur. As we will demonstrate, however, this does not imply that the principal is better off appointing an expert.

6.1 Preliminary intuitions

It is expositionally useful to postpone the statement of our general results and begin with an intuitive presentation of the key mechanisms at work. Specifically, there are two reasons why the principal may be better off appointing an amateur rather than the expert.

The first reason is that an agent's expertise may force the principal to offer her too

[18] Lewis (2007) and Kelman and Myers (2009) provide evidence which supports this conjecture, even though civil servants may be unduly biased towards the status quo (cf. McCarty (2004)).

[19] By contrast, Bendor and Meirowitz (2004) and Egorov and Sonin (forthcoming) model an amateur as an agent who knows the state with some probability and otherwise has the same prior beliefs as the principal.

many decisions. To see this, it is useful to focus on putative improvements which replace a subset of decisions in the expert's ODS. Specifically, suppose that there are four states, and that $1/3 < b < 1/2$. The ODS, which we describe in Appendix C.2, is

$$\Delta = \left\{ \frac{1}{2}b, \frac{3}{2}b, \frac{1}{2}b + \frac{2}{3} \right\} ,$$

and the principal loses $\frac{5}{4}b^2 - \frac{1}{2}b + \frac{1}{18}$ in event $E \equiv \{0, 1\}$. Consider another delegation set $\Delta' = \{\frac{3}{2}b - \frac{1}{3}, \frac{1}{2}b + \frac{2}{3}\}$: where $\frac{3}{2}b - \frac{1}{3}$ replaces the lower two decisions in Δ. If the agent were always to take $\frac{3}{2}b - \frac{1}{3}$ in E then the principal would lose $\frac{9}{4}b^2 - \frac{3}{2}b + \frac{5}{18}$ in E, which is less than $\frac{5}{4}b^2 - \frac{1}{2}b + \frac{1}{18}$. In other words, the principal would lose less when he offers Δ', provided that the agent takes $\frac{3}{2}b - \frac{1}{3}$ in E. The latter condition fails when the agent is an expert because she would take $\frac{1}{2}b + \frac{2}{3}$ in state 1.

Now consider an amateur who cannot distinguish between the states in E, but knows the exact state otherwise. The first decision in Δ' has been calibrated such that the amateur is indifferent across Δ' when the state is in E, and strictly prefers $\frac{1}{2}b + \frac{2}{3}$ in states 2 and 3 (by the single-crossing property). Consequently, this amateur would always take $\frac{3}{2}b - \frac{1}{3}$ in E, and would take $\frac{1}{2}b + \frac{2}{3}$ in states 2 and 3. In sum, the principal is better off appointing this amateur and offering her Δ' than appointing the expert and offering her ODS.

Notice that we have constructed Δ' such that the principal only gains in those states that the amateur cannot distinguish. Most of our arguments below will rely on this property; but it is important to understand that the principal might also gain in states outside E: the second reason why an amateur might outperform the expert. To see this, note that the decisions taken in states 2 and 3 are also too high. The principal could therefore improve on Δ' by offering another pair of decisions which satisfy the chain property (the agent is indifferent between them in event E). Calculations in Appendix C.2 reveal that the ODS for this amateur is $\{b - \frac{1}{6}, b + \frac{1}{2}\}$: the principal improves on Δ' by raising the first and reducing the second decision in Δ'. This reinforces our observation that the principal prefers to appoint an amateur than the expert. Indeed, it turns out that the principal cannot improve on appointing this amateur.

By construction, agents share the same bias, and therefore make the same loss after any history, irrespective of whether they have been appointed to take the decisions. It is easy to confirm that the agents prefer the principal to offer the expert her ODS than to offer Δ' to an amateur, but that they are better off if the principal appoints the amateur than the expert if he would offer each agent her ODS. However, agents would be best off if the principal appointed an agent who cannot distinguish between states 1 and 2.

6.2 Agent performance

We now turn to the more general setting of T states, again supposing that the principal can appoint the expert or an agent who cannot distinguish between states in some specified events E (i.e., an *amateur*). We will say that some amateur *outperforms* the expert if the principal loses more by appointing the expert than appointing some amateur if he offers each agent her loss-minimizing delegation set (ODS). Our main result in this subsection states that (i) the expert cannot be outperformed by any amateur when b is high enough that delegation is not valuable or when b is low enough that the principal can achieve first best with the expert; but (ii) for intermediate bias, there is an amateur who outperforms the expert.

Proposition 3. ***(i)*** *The expert cannot be outperformed by any amateur if* $b \leq \frac{1}{2(T-1)}$ *or* $b > \frac{T}{2(T-1)}$.

(ii) *Some amateur outperforms the expert if there are at least four states and* $\frac{1}{T-1} < b < \frac{T}{2(T-1)}$.

We prove Proposition 3 in Appendix B.

Proposition 3 generalizes our arguments in the last subsection. In particular, it asserts that some amateur outperforms the expert for intermediate bias. Our construction of a superior amateur generalizes that used in the last subsection when $K > 1$: such an amateur cannot distinguish between states $K-2$ and $K-1$; and we again construct a delegation set such that the amateur and the expert take the same decisions in all states other than $K-2$ and $K-1$. Our construction differs when $b \in (\frac{1}{2}, \frac{T}{2(T-1)})$, when the expert would take two decisions: the superior amateur can now not distinguish between states 0 and 1. In that event, she takes a decision below that taken on average by the expert, and takes a higher decision than the expert in states $t > 1$. As above, we prove the result by showing that the principal loses less by offering some delegation set to some amateur, rather than by characterizing a given amateur's ODS.

The premise of part (ii) excludes $b \in (\frac{1}{2(T-1)}, \frac{1}{T-1})$: when the ODS contains T decisions. It is easy to confirm that the expert outperforms an amateur who cannot distinguish between states $T-3$ and $T-2$, and takes the same decisions as the expert in all other states.[20] This observation and part (ii) imply that no amateur outperforms the expert if there are fewer than four states. This is obvious when $T = 2$, and slightly more subtle when $T = 3$. Arguments used in the proof of part (ii) then imply that no amateur can outperform the expert if $b > 1/(T-1)$; the observation implies that an amateur who cannot distinguish between states 0 and 1 cannot outperform the expert; and an argument at the end of the last subsection precludes an amateur who cannot distinguish between states 1 and 2 outperforming the expert.[21]

As T increases, the interval $(\frac{1}{2(T-1)}, \frac{1}{T-1})$ shrinks; so, for arbitrarily large T, an amateur outperforms the expert whenever delegation is valuable. This is consistent with Ivanov (2010) Theorem 4, which asserts that some amateur outperforms the expert in Holmstrom's model whenever delegation is valuable. Ivanov attributes this result to the fact that an expert then takes her ideal decision in some states (p736). Proposition 3 shows that some amateur outperforms the expert even if the principal compromises with the expert: a feature of the ODS.

Proposition 3 emphasizes the significance of our approach to modelling 'amateurs'. We have supposed that an amateur cannot distinguish between some states, whereas Bendor

[20] We provide a proof of this claim, which we dub Observation 3, in the online Appendix (Anesi and Seidmann (2011)).

[21] Delegation to an amateur who cannot distinguish between states 0 and 2 is never valuable because her preferences over decisions are state-independent.

and Meirowitz (2004) model an amateur as an agent who shares the principal's beliefs with positive probability, and otherwise observes the state exactly. Bendor and Meirowitz note that the principal would never appoint an amateur of their sort unless she is less biased than the expert. By contrast, Proposition 3 explains evidence of incompetent political appointees without reference to a tradeoff betweeen loyalty and expertise. Indeed, it implies that the principal could optimally appoint an agent who is less loyal than an expert.

6.3 Preferences over expertise

Thus far, we have focused on the principal's point of view, as in Ivanov (2010). In contrast to the previous literature, we now consider the agent's preferences over her expertise. Specifically, we imagine that the agent can costlessly choose a possibly empty set of events such that she cannot distinguish between states in each event. The principal observes the agent's expertise and then offers her the delegation set which minimizes the principal's loss (viz. the ODS).[22]

To be concrete, consider the relationship between a defendant (principal) and his attorney (agent). The attorney discovers the state via conversations with the defendant, who instructs the attorney on how to conduct the case (the delegation set). The attorney shares the defendant's interests, but is required (by ABA ethics rules) to disclose his intended perjury to the court:[23] an effect analogous to a bias. Accordingly, the attorney would not ask questions whose answers might require disclosure: an instance of willful ignorance. In our terms, the attorney commits not to observe the exact state.

We will argue that the agent in our model would choose not to become the expert for some bias levels. The driving force for this result turns on the degree to which the two players share preferences over decision-state pairs. In particular, we do not assume that greater expertise is more costly. In other words, there is no trade-off between cost and expertise in our model.

It is obvious that the agent would choose to be the expert when b is low enough: for the expert's ODS then induces the agent to take decisions which are almost ideal for her in every state; whereas the decision which she takes would not depend on the state in some event(s) if she chose to be an amateur. In sum, the agent shares the principal's preference for an expert when b is low enough (cf. Proposition 3).

It is also obvious that the principal and the agent are indifferent across expertise when b is so high that delegation is not valuable. More interestingly, our next result shows that the principal and the agent both prefer that the agent be some amateur rather than the expert when $b \in (\frac{1}{2}, \frac{T}{2(T-1)})$: that is, the expert's ODS consists of two decisions (cf. Theorem 1):

Proposition 4. *If $b \in (\frac{1}{2}, \frac{T}{2(T-1)})$ then the principal and the agent are both better off when the agent cannot distinguish between states 0 and 1 than when she is the expert.*

[22] An equivalent interpretation is that agents with diverse expertise choose whether to become candidates for appointment.

[23] See ABA Model Rules of Professional Conduct Rule 3.3(b).

We prove Proposition 4 in Appendix B. The proof demonstrates that the ODS for this amateur ($\{\delta_0^{01}, \delta_1^{01}\}$) consists of two decisions for this range of biases, and that the agent takes the lower decision in states 0 and 1 alone. We can then calculate the principal's and the agent's losses, yielding the result.

The intuition for the agent's preferences can best be seen for b close to $T/2(T-1)$ when the lower decision in each ODS is close to $1/2$.[24] On the other hand, δ_1^* and δ_1^{01} satisfy

$$\delta_1^* \simeq \frac{1}{T-1} + \frac{1}{2} < \frac{2}{T-1} + \frac{1}{2} \simeq \delta_1^{01}.$$

It is easy to confirm that the expert loses less than the amateur in state 1. However, δ_1^{01} is less than the agent's ideal decision in states t, so the expert's gain in state 1 is over-ridden by the extra losses in higher states.

Proposition 4 is consistent with anecdotal evidence that defendants and attorneys both prefer that the attorney not ask questions which are too detailed.

Proposition 4 can also be read from a normative standpoint. A regulator concerned with the players' welfare who could only choose an information structure would not necessarily require that the agent be an expert.

Proposition 4 and the preceding discussion imply that the agent prefers to be the amateur for some, but not all bias levels. This result parallels Luban's (1999) argument that the ethics of the attorney's willful ignorance turn on her motives.

The principal's and the agent's preferences over expertise are more complicated when b is intermediate. Indeed, to the best of our knowledge, the identification of the best amateur, from either player's point of view, remains an unsolved problem in the Holmstrom's continuous-state model: probably because of the technical difficulties one would have to overcome. Identifying the principal's best amateur is also a challenging problem in our model because the ODS for a given amateur need not satisfy either the chain property or top-loading. Nevertheless, we can make some progress by exploiting our previous results; in particular:

- The ODS for amateur E (ODS_E) shares the recursive structure of the ODS for an expert. Thus, ODS_E minimizes the expected loss for states outside E subject to incentive constraints for states in $T \backslash E$ and for states just below E. The constraints in $T \backslash E$ are also faced by the expert; so we can use Lemma 2 to characterize the decisions taken for states in $T \backslash E$.

- Proposition 3 establishes a benchmark when $K < T-1$: the best amateur must outperform the expert. This immediately implies that amateur who cannot disitnguish between states in $E \equiv \{t, ..., T-1\}$ cannot be best, as the principal can offer ODS_E to the expert.

Combined with these observations, the finiteness of the set of states allows us to calculate the principal's and the agent's ideal expertise for cases with less than six states. (We will treat *ideal expertise* and *best amateur* as synonyms below.) We will simplify notation

[24] See the proof of Proposition 4 for the amateur's ODS, and Section 3.2 for the expert's ODS.

by denoting an agent who cannot distinguish states s and t as agent st, an agent who cannot distinguish between s and t or between u and v as st, uv, etc.

It is easy to see that cases with two or three states are uninteresting: delegation is not valuable for the only amateur in the former case; while amateurs 01 and 12 are payoff-equivalent (and no better than the expert) when there are three states. We calculate the ODS for the expert and for amateurs when there are four or five states in Appendices C.2 and C.3.[25] We summarize the expertise that the principal and the agent would choose for every bias

Table 1 describes the principal's and the agent's ideal expertise when $T = 4$. The expert's ODS comprises four decisions ($K = 3$) when $b < 1/3$, three decisions ($K = 2$) when $b \in (1/3, 1/2)$, and two decisions ($K = 1$) when $b \in (1/2, 2/3)$.

b	Principal's ideal expertise	Agent's ideal expertise
$b < \frac{7}{24}$	Expert	Expert
$\frac{7}{24} < b < \frac{1}{3}$	12	Expert
$\frac{1}{3} < b < \frac{3}{8}$	12	12
$\frac{3}{8} < b < \frac{1}{2}$	01	12
$\frac{1}{2} < b < \frac{2}{3}$	01	01
$b > \frac{2}{3}$	All	All

Table 1: $T = 4$

If $b > 5/8$ then delegation to an expert is not valuable, so it is not valuable to any other agent, all of whom are therefore (trivially) best for both players.

If $3/8 < b < 2/3$ then the principal would offer the same ODS to 01 and to an amateur who can also not distinguish between states 2 and 3. The two amateurs are equivalent because 01 takes the same decisions in states 2 and 3 according to ODS_{01}. The best amateur is also not uniquely defined when $b < 1/6$: all pairs are payoff-equivalent (and are outperformed by the expert).

Proposition 3 asserts that the expert outperforms any amateur when $b \leq 1/2(T-1)$, and therefore whenever b just exceeds $1/2(T-1)$. Table 1 reveals that an amateur who cannot distinguish between states $K-2$ and $K-1$ outperforms the expert when $b \in (b^+, 1/(T-1))$ for some $b^+ > 1/2(T-1)$. This amateur corresponds to that used in our proof of Proposition 3(ii). However, contrary to that construction, she takes different decisions than the expert in states 0 and 3 according to her ODS. This property confirms the importance of the second reason for appointing an amateur, as detailed in Section 6.1

[25] In Appendix C, we only describe details for amateurs who appear in our discussion below. We report details for the omitted amateurs in Anesi and Seidmann (2011).

above: the principal may relax the incentive constraints of an amateur in states outside E by changing the decision which she takes in E. On the other hand, the best amateur cannot distinguish between $K-1$ and K when $b > 1/(T-1) = 1/3$: contrary to the construction which we used to prove Proposition 3(ii).[26] Furthermore, the principal offers fewer decisions to the best amateur than to the expert when $b \in (\frac{7}{24}, \frac{1}{3}) \cup (\frac{3}{8}, \frac{1}{2})$.

Calculations in Appendix C.2 also reveal that the ODS for the principal's ideal expertise always satisfies the chain property: generalizing the characteristics of the expert's ODS. In addition, the expert is never outperformed by an amateur who cannot distinguish more than two states. These calculations also reveal that the agent is indifferent between being an expert and some amateur (01) if $b < 1/3$, and prefers to be an amateur (12) if $1/3 < b < 2/3$.

Table 2 provides equivalent details for the five-state case, where the expert's ODS comprises five decisions when $b < 1/4$, four decisions when $b \in (1/4, 3/8)$, three decisions when $b \in (3/8, 1/2)$, and two decisions when $b \in (1/2, 5/8)$.

Value of b	Principal's ideal expertise	Agent's ideal expertise
$b < \frac{1}{4}$	Expert	Expert
$\frac{1}{4} < b < \frac{9}{32}$	12	Expert
$\frac{9}{32} < b < \frac{1}{3}$	23	Expert
$\frac{1}{3} < b < \frac{3}{8}$	12	Expert
$\frac{3}{8} < b < \frac{1}{2}$	12	12 and 01, 23
$\frac{1}{2} < b < \frac{11+\sqrt{265}}{48}$	01	12 and 01, 23
$\frac{11+\sqrt{265}}{48} < b < \frac{5}{8}$	01	01 and 02, 14 and 012
$b > \frac{5}{8}$	All	All

Table 2: $T = 5$

Table 2 reveals various differences from the four-state case. The principal's best amateur outperforms the expert for some $b < 1/(T-1)$ when $T = 4$; Table 2 shows that this is impossible when $T = 5$.

If the principal's best amateur outperforms the expert then she cannot distinguish between states $K-2$ and $K-1$ or between $K-1$ and K, as in the $T = 4$ case; but in contrast, E switches between $(K-2)(K-1)$ and $(K-1)K$ for this range of bias. This non-monotonicity seems to occur because of the interplay of two effects. If b is low enough then E is low: for replacing the expert with amateur $t(t+1)$ relaxes the incentive

[26] She again takes different decisions than the expert in states 0 and 3 when offered her ODS.

constraint in state $t+1$, allowing the agent to take lower decisions in every state $\tau \geq t+2$, which is advantageous to the principal because the expert takes decisions above $\tau/(T-1)$ in each such state (cf. Corollary 1). On the other hand, the principal incurs extra losses in some states $\tau < t+2$; so the best amateur is not driven down to 01 when $T = 5$. Two effects occur as b increases. The number of decisions in the expert's ODS for states $\tau \geq t+2$ falls, reducing the gain from relaxing incentive constraints in those states. On the other hand, an amateur who cannot distinguish states above an expert's K cannot outperform an expert; so there is a countervailing effect which starts to reduce the best E. This non-monotonicity also extends to the number of decisions that the best amateur takes: if $\frac{1}{4} < b < \frac{9}{32}$ then 12 takes three decisions; and if $\frac{9}{32} < b < \frac{1}{3}$ then 23 takes four decisions. (See Appendix C.3 for details.)

When $T = 4$ the ODS for the principal's best agent satisfies the chain property. This is no longer true with five states. Specifically, if $1/4 < b < 9/32$ then the principal's best amateur (12) strictly prefers the decision she takes in state 3 over the single decision which she takes in states 3 and 4.[27] The analogous ODS also fails the chain property when $9/32 < b < 1/3$.

While the principal's best amateur cannot distinguish pairs of successive states, amateur 123, who cannot distinguish three successive states, outperforms the expert when $5/16 < b < 1/3$. We conjecture that the principal's best amateur may be unable to distinguish more than two states when T is large enough.

When $T = 4$, the ODS offered to each player's ideal expertise is uniquely defined for generic bias. This property does not carry over when $T = 5$: various amateurs yield the agent her minimal loss, and each of these agents is offered a different ODS. In further contrast to the four state case, some of these agents cannot distinguish between more than a pair of neighboring states.

We have noted above that the principal and agent top rank the same agent when b is very low and when it is so high that delegation is not valuable. Tables 1 and 2 demonstrate that this property carries over to intermediate bias. This strengthens Proposition 4 in two senses: 01 is best for both players as well as prefered to the expert; and the same agent is best for both players when $b < 1/2$ and $T = 5$.

If there are four or five states then the agent prefers to be the expert for biases at which the principal would appoint an amateur. The intuition is as follows. At the critical b where the principal is indifferent, the ODS for the amateur consists of fewer decisions than the ODS for the expert, who takes a decision which is too high for the principal in every state (cf. Corollary 1(i)). Reducing state-sensitivity is, ceteris paribus, bad for the principal; so the amateur must take lower decisions in some states. As the expert's decisions are too low for the agent (cf. Corollary 1(ii)), she must be worse off when the principal appoints the amateur.

[27] Consequently, the latter decision is first best for the principal in those states.

7 Conclusion

A growing literature has built on Holmstrom's (1984) uniform-quadratic model to explore the decisions induced by an optimal mechanism without money transfers when the support of states is an interval. The ODS in Holmstrom (1984) has a simple structure, which generates testable implications, like the Ally Principle. The literature has provided necessary and sufficient conditions on preferences and the distribution of states for implications like the Ally Principle.

Interval support alone implies that the ODS has a no-compromise property: the agent must take her ideal decision in those states where her reaction function is locally increasing (cf. Melumad and Shibano (1991)). We have extended the literature by studying the implications of finite support: both because the support of states is finite (and sometimes small) in empirically salient situations, and because of both technical and testable implications.

We have focused on the simplest extension of Holmstrom's model: to emphasize the implications of finite support in an otherwise well-understood structure; and, like various papers in the literature, to exploit its tractability.

We have demonstrated that an expert's ODS has a strikingly simple form, which is characterized by two properties: top loading implies that the agent is indifferent in low enough states; while the chain property implies that the decision taken in those states compromises between the principal's and the agent's ideal decisions. The chain property implies that local variations in the agent's bias shift the ODS uniformly upwards: a testable feature which we have contrasted with the Ally Principle. We have also exploited the simple structure of the ODS to apply our model to the choice of agents, demonstrating that the Expertise Principle fails for intermediate bias, and characterizing the principal's and the agent's best agents when there are few states.

The chain property and top loading rely on quadratic losses and equi-probable and equi-spaced states. To see this, consider a generalization in which the three, equi-probable states are 0, ε and 2, where $\varepsilon < 4/7$; and players incur the same losses as in our model. We describe the expert's ODS for generic bias in Table 3 below:

	Δ^*
$b < \frac{\varepsilon}{4}$	$0, \frac{\varepsilon}{2}, 1$
$\frac{\varepsilon}{4} < b < \frac{\varepsilon}{2}$	$b - \frac{\varepsilon}{4}, b + \frac{\varepsilon}{4}, 1$
$\frac{\varepsilon}{2} < b < \frac{4-3\varepsilon}{8}$	$\frac{\varepsilon}{4}, 1$
$\frac{4-3\varepsilon}{8} < b < \frac{4-3\varepsilon+\sqrt{6\varepsilon^2-18\varepsilon+36}}{8}$	$\frac{2}{3}b - \frac{2-3\varepsilon}{6}, \frac{4}{3}b + \frac{2+3\varepsilon}{6}$
$b > \frac{4-3\varepsilon+\sqrt{6\varepsilon^2-18\varepsilon+36}}{8}$	$\frac{2+\varepsilon}{6}$

Table 3

The chain property fails for $\frac{\varepsilon}{4} < b < \frac{4-3\varepsilon}{8}$, when the agent takes the principal's ideal decision in some, but not all states. The agent only takes the same decision in states ε and 1 when delegation is not valuable; and she takes the same decision in states 0 and ε when

$$b \in (\frac{\varepsilon}{2}, \frac{4 - 3\varepsilon + \sqrt{6\varepsilon^2 - 18\varepsilon + 36}}{8}):$$

so top loading fails.

While the ODS does not satisfy either the chain property or top loading, the agent never takes her ideal decision; and any decision which is not ideal for the principal is increasing in b. Consequently, any generic marginal increase in bias raises some decisions and leaves the other decisions fixed. In other words, comparative statics with respect to the bias are similar to those in our model. We conjecture that these features generalize to other cases with a finite support, upward biased agent and preferences which satisfy single crossing: the agent never takes her ideal decision because the principal can exploit the distance between the agent's ideal decisions across states; and an increase in the bias never reduces the decision taken in any state.

To see the latter, let b_0 be just below b_1, and write $\{d_\tau(b_i)\}$ for the ODS when the bias is b_i $(i = 0, 1)$. Suppose, contrary to our claim, that there is a state t such that $d_t(b_1) < d_t(b_0)$. Optimality of both ODSs then implies that there is a state $s < t$ at which the agent with the lower bias strictly prefers $d_t(b_1)$ over $d_s(b_0)$; so single crossing implies that $d_s(b_0) < d_s(b_1)$. Optimality of $\{d_\tau(b_1)\}$ then implies that the principal can profitably deviate when the bias is b_0.

APPENDIX A: Proof of Theorem 1

Proof of Lemma 1

Without loss of generality, we normalize t to zero. Define the optimization problem $(\mathcal{P})$ as

$$\min_{(d_0,\ldots,d_{s-1})} \sum_{i=0}^{s-1} \left(d_i - \frac{i}{T-1} \right)^2 ,$$

$$\text{subject to } \left(d_i - b - \frac{i}{T-1} \right)^2 \leq \left(d_{i+1} - b - \frac{i}{T-1} \right)^2 , \forall i = 0, \ldots, s-1.$$

Our first step is to establish the existence of a solution to problem $(\mathcal{P})$ by using the Weierstrass Theorem. To do so, we must ensure that the objective function in $(\mathcal{P})$ is continuous, and that the collection of incentive-compatibility constraints defines a feasible set that is compact in $\mathbb{R}^s$. While continuity of the principal's loss function and closedness of the feasible set are trivial (inequalities are weak and the agent's loss function is always continuous), the feasible set of $(\mathcal{P})$ is obviously not bounded.

Nevertheless, any solution to $(\mathcal{P})$, say $\left(d_0^*, \ldots, d_{s-1}^*\right)$, must satisfy

$$0 \leq d_i^* \leq 2 \left(b + \frac{s-2}{T-1} \right) , \tag{6}$$

for each $i = 0, \ldots, s-1$. To see this, suppose first that some of the d_i^*'s are negative. In such a situation, the principal can reduce her loss by replacing every $d_i^* < 0$ by $d_i = 0$. Indeed, the agent would take decision 0 in every state $i \leq I \equiv \max\{i : d_i^* < 0\}$, making the principal strictly better off. Moreover, the principal would also be better off in every state $i > I$ in which the agent would take 0 instead of $d_i^* > 0$. This proves that $d_i^* \geq 0$, for each $i = 0, \ldots, s-1$.

Furthermore, the incentive-compatibility constraint in state $i+1$, $i < s-1$, requires that $d_i^* \leq b + \frac{s-1}{T-1}$ whenever $d_i^* < d_{i+1}^*$. Now, to show that the above inequality is also true for $i = s-1$, let $\tau + 1 \equiv \min\left\{j : d_j^* = d_{s-1}^*\right\}$ and suppose that $d_{s-1}^* > b + \frac{s-1}{T-1}$. This implies that the incentive-compatibility constraint in state τ must be satisfied with equality, else the principal could reduce his expected loss by reducing $d_{\tau+1}^*$. Thus, either $d_\tau^* = d_{\tau+1}^*$ or

$$d_\tau^* + d_{\tau+1}^* = 2 \left(b + \frac{\tau}{T-1} \right) .$$

Combined with $d_\tau^* \geq 0$, this implies that

$$d_i^* \leq d_{s-1} = d_{\tau+1}^* \leq \max\left\{ b + \frac{s-1}{T-1}, 2\left(b + \frac{\tau}{T-1} \right) \right\} \leq 2 \left(b + \frac{s-1}{T-1} \right)$$

for each $i = 0, \ldots, s-1$.

We have thus established that any solution to $(\mathcal{P})$, $\left(d_0^*, \ldots, d_{s-1}^*\right)$, must satisfy (6). Consequently, problem $(\mathcal{P}')$, obtained by adding the constraints (6) to $(\mathcal{P})$, has the same solution(s) as $(\mathcal{P})$. The conditions in (6) guarantee that the feasible set of $(\mathcal{P}')$ is bounded (and therefore compact). By Weierstrass Theorem, $(\mathcal{P}')$ – and therefore $(\mathcal{P})$ – has at least one solution.

Let $\left(d_0^*, \ldots, d_{s-1}^*\right)$ be a solution to $(\mathcal{P})$. Recursively define

$$\begin{aligned} \delta_0^* &\equiv d_0^* \ , \\ D_i &\equiv \left\{d \in \left\{d_0^*, \ldots, d_{s-1}^*\right\} : d > \delta_{i-1}^*\right\} \ , \\ \delta_i^* &\equiv \min_{d \in D_i} d \ . \end{aligned}$$

If K is the smallest integer such that $D_{K+1} = \emptyset$ then $\{\delta_0^*, \ldots, \delta_K^*\}$ is, by definition, an optimal delegation set.

Proof of Lemma 2

As Δ is top loaded, the agent takes decision $\delta_{t+\tau}$ in state $t+\tau$, for every $\tau < K$, and takes δ_{t+K} in states $K \leq t+\tau \leq s-1$.

For each $\tau \leq K$, let $\kappa_{t+\tau} \equiv \delta_{t+\tau} - \frac{t+\tau}{T-1}$. By optimality, $(\kappa_{t+\tau})_{\tau=0\ldots,K}$ must solve

$$\min_{\{\kappa_{t+\tau}\}_{\tau=t,\ldots,K}} \left\{\sum_{\tau=0}^{K-1} \kappa_{t+\tau}^2 + \sum_{\tau=K}^{s-1} \left(\kappa_{t+K} - \frac{\tau-K}{T-1}\right)^2\right\}$$

subject to

$$\kappa_{t+\tau} + \kappa_{t+\tau+1} = 2b - \frac{1}{T-1} \ , \forall \tau = 0, \ldots, K-1 \ ; \tag{7}$$

where the equality in (7) must hold at Δ because it satisfies the chain property. It is easy to see that any solution to the above problem is also a solution to

$$\min_{\{\kappa_{t+\tau}\}_{\tau=t,\ldots,K}} \left\{\sum_{\tau=0}^{K-1} \kappa_{t+\tau}^2 + (s-K)\left(\kappa_{t+K} - \frac{s-K-1}{T-1}\right)\kappa_{t+K}\right\} \ , \tag{8}$$

subject to (7). From the incentive constraints (7), we obtain by recursion that

$$\kappa_{t+\tau} = \begin{cases} \kappa_t & \text{if } \tau \text{ is even} \\ 2b - \frac{1}{T-1} - \kappa_t & \text{if } \tau \text{ is odd} \end{cases} \ , \forall \tau = 0, \ldots, K.$$

Substituting into (8), simple convex optimization reveals that, at an optimum:

$$\kappa_t = \begin{cases} \ell^e(b,s,K) & \text{if } K \text{ is even,} \\ \ell^o(b,s,K) & \text{if } K \text{ is odd,} \end{cases}$$

where $\ell^e(b,s,K)$ and $\ell^o(b,s,K)$ are defined before Lemma 2 in Section 2.3.

We end the proof by showing that the principal can reduce his expected loss whenever $b \notin \left[b^{\min}(K,s), b^{\max}(K,s)\right]$ — a contradiction to Δ being optimal. More precisely, for every $m \in \mathbb{N}$, let $D^*(m) \equiv \left\{\delta_t^*(m), \ldots, \delta_{t+m-1}^*(m)\right\}$ where

$$\delta_{t+\tau}^*(m) = \begin{cases} \ell(b,s,m-1) + \frac{t+\tau}{T-1} & \text{if } \tau \text{ is even,} \\ 2b - \ell(b,s,m-1) + \frac{t+\tau-1}{T-1} & \text{if } \tau \text{ is odd.} \end{cases}$$

We will show that, if $b \notin \left[b^{\min}(K,s), b^{\max}(K,s)\right]$ then there is $K' \neq K$ such that the principal can reduce his loss by offering either $D^*(K'-1)$ or $D^*(K'+1)$ to the agent.

Suppose that K is even. The change in the principal's loss following a deviation to $D^*(K+2)$ when $K < s-1$ can be decomposed as follows:

(i) In each state $t+\tau$, with τ even and no more than K, the change in the principal's loss is:

$$\ell^o(b,s,K+1)^2 - \ell^e(b,s,K)^2 = \varphi(b,s,K)\left[\varphi(b,s,K) + 2\ell^e(b,s,K)\right] ,$$

where

$$\varphi(b,s,K) \equiv \ell^o(b,s,K+1) - \ell^e(b,s,K) = \frac{s-K-1}{s}\left(2b - \frac{s-K}{T-1}\right) .$$

(ii) In each state $t+\tau$, with τ odd and no more than $K-1$, the change in the principal's loss is:

$$\left[2b - \frac{1}{T-1} - \ell^o(b,s,K+1)\right]^2 - \left[2b - \frac{1}{T-1} - \ell^e(b,s,K)\right]^2$$
$$= \varphi(b,s,K)\left[\varphi(b,s,K) - 2\left(2b - \frac{1}{T-1} - \ell^e(b,s,K)\right)\right] .$$

(iii) In each state $t+K+i$, $i = 1,\ldots,s-K-1$, the change in the principal's loss is:

$$\left[2b - \ell^o(b,s,K+1) - \frac{i}{T-1}\right]^2 - \left[\ell^e(b,s,K) - \frac{i}{T-1}\right]^2$$
$$= \xi(b,s,K)\left\{\xi(b,s,K) + 2\left[\ell^e(b,s,K) - \frac{i}{T-1}\right]\right\} ,$$

where

$$\xi(b,s,K) \equiv 2b - \ell^o(b,s,K+1) - \ell^e(b,s,K) = \frac{1}{s}\left(2b + \frac{K}{T-1}\right) .$$

Note that

$$\sum_{i=1}^{s-K-1}\left\{\xi(b,s,K) + 2\left[\ell^e(b,s,K) - \frac{i}{T-1}\right]\right\} = (s-K-1)\left[\xi(b,s,K) + 2\ell^e(b,s,K) - \frac{s-K}{T-1}\right]$$
$$= (K+1)\frac{s-K-1}{s}\left(2b - \frac{s-K}{T-1}\right)$$
$$= (K+1)\varphi(b,s,K) .$$

Thus, summing the changes in the principal's loss over all states, we obtain

$$(K+1)\varphi(b,s,K)\left[\varphi(b,s,K) + 2\ell^e(b,s,K) - \frac{K}{K+1}\left(2b - \frac{1}{T-1}\right) + \xi(b,s,K)\right]$$

$$= (K+1)\varphi(b,s,K)\left[\frac{1}{K+1}\left(2b + \frac{K}{T-1}\right)\right] = \frac{2(s-K-1)}{s}\left(2b + \frac{K}{T-1}\right)\left[b - b^{\min}(K,s)\right] .$$

This proves that the deviation to $D^*(K+2)$ is unprofitable only if $b \geq b^{\min}(K,s)$. The argument above also implies that the principal can profitably deviate from $D^*(K+2)$ to $D^*(K+1)$ unless $b \leq b^{\max}(K+1,s) = b^{\min}(K,s)$.

Analogous arguments imply (see the online Appendix for a detailed proof) that the change in the principal's expected loss if he deviates to $D^*(K)$ (when $K > 0$) equals

$$-\frac{2K}{T-1}\frac{s-K}{s}\left[b - b^{\max}(K,s)\right] \ .$$

This proves that the deviation to $D^*(K)$ is unprofitable only if $b \leq b^{\max}(K,s)$. The above argument also implies that the principal can profitably deviate from $D^*(K)$ to $D^*(K+1)$ unless $b \geq b^{\min}(K-1,s) = b^{\max}(K,s)$.

We have therefore proved that the principal can reduce his expected loss by choosing $D^*(K+2)$ instead of Δ when $b < b^{\min}(K,s)$, and by choosing $D^*(K)$ instead of Δ when $b > b^{\max}(K,s)$. [28] As $\Delta \in \mathcal{D}^*_{s,t}(b)$, this implies that $b \in \left[b^{\min}(K,s), b^{\max}(K,s)\right]$.

Proof of Lemma 3

Suppose that $(\mathbf{H_s})$ is true and that, contrary to the Lemma, some element of $\mathcal{D}^*_{s+1,t}(b)$, say Δ^*, does not satisfy the chain property.

$(\mathbf{H_s})$ and the optimality of Δ^* imply that the chain can only break once. To see this, note that there would otherwise be k_1 and k_2 such that the agent would strictly prefer $d_{t+\tau_i}(\Delta^*, b)$ to $d_{t+\tau_i+1}(\Delta^*, b)$, $i = 1,2$ in states $t + \tau_i = t_{k_i}(\Delta^*, b)$, $t \leq \tau_1 < \tau_2 \leq t+s$. A brief inspection of problem of the principal's optimization problem (see above) reveals that this would in turn imply that $\Delta^* = \Delta_1 \cup \Delta_2 \cup \Delta_3$, where $\Delta_1 \in \mathcal{D}^*_{\tau_1+1,t}(b)$, $\Delta_2 \in \mathcal{D}^*_{\tau_2-\tau_1,t+\tau_1+1}(b)$, and $\Delta_3 \in \mathcal{D}^*_{s-1-\tau_2,\tau_2+1}(b)$. But this would contradict $\Delta^* \in \mathcal{D}^*_{s+1,t}(b)$: for $(\mathbf{H_{\tau_2+1}})$ implies that $(\Delta_1 \cup \Delta_2) \notin \mathcal{D}^*_{\tau_2+1,t}(b)$; so the principal could reduce his expected loss by replacing Δ^* with $(\Delta_1' \cup \Delta_3)$, where $\Delta_1' \in \mathcal{D}^*_{\tau_2+1,t}(b)$.

In sum, there is a unique k such that the agent strictly prefers $d_{t+\tau-1}(\Delta^*, b)$ over $d_{t+\tau}(\Delta^*, b)$ in state $t+\tau-1 = t_k(\Delta^*, b) \geq 1$. Consequently, there is $\Delta_1^* \in \mathcal{D}^*_{\tau,t}(b)$ and $\Delta_2^* \in \mathcal{D}^*_{s-\tau+1,t+\tau}(b)$ such that $\Delta^* = \Delta_1^* \cup \Delta_2^*$, and

$$d_{t+\tau-1}\left(\Delta_1^*, b\right) + d_{t+\tau}\left(\Delta_2^*, b\right) = d_{t+\tau-1}(\Delta^*, b) + d_{t+\tau}(\Delta^*, b) > 2\left(b + \frac{t+\tau-1}{T-1}\right) \ . \quad (9)$$

Lemma 2 implies that there are integers $K \leq \tau - 1$ and $K' \leq s - \tau$ such that

$$d_{t+\tau-1}\left(\Delta_1^*, b\right) = \begin{cases} \ell^e(b,\tau,K) + \frac{t+K}{T-1} & \text{if } K \text{ is even,} \\ 2b - \frac{1}{T-1} - \ell^o(b,\tau,K) + \frac{t+K}{T-1} & \text{if } K \text{ is odd,} \end{cases} \ , \quad (10)$$

$$d_{t+\tau}\left(\Delta_2^*, b\right) = \begin{cases} \ell^e(b, s-\tau+1, K') + \frac{t+\tau}{T-1} & \text{if } K' \text{ is even} \\ \ell^o(b, s-\tau+1, K') + \frac{t+\tau}{T-1} & \text{if } K' \text{ is odd} \end{cases} \ , \quad (11)$$

and

$$b \geq \max\left\{\frac{\tau - K}{2(T-1)}, \frac{s-\tau+1-K'}{2(T-1)}\right\} \ . \quad (12)$$

[28] Note that we have ignored the case in which $K = s-1$ and $b < b^{\min}(K,s)$, since it implies $b < 1/2(T-1)$.

To obtain the desired contradiction, we will use (12) to prove that (10) and (11) are inconsistent with (9). Before we proceed any further, however, the following observations are worth making. First, $b \geq 1/2(T-1)$ implies that, for any s and any K,

$$2b - \frac{1}{T-1} - \ell^o(b,s,K) - \ell^e(b,s,K) = \frac{1}{s}\left(b - \frac{1}{2(T-1)}\right) \geq 0 \ .$$

Second, $b \geq (s-K)/2(T-1)$ implies that, for any s and any K,

$$\ell^e\left(b,s,K\right) - \ell^o\left(b,s,K\right) \geq \frac{2(K-s)+1}{s}\frac{s-K}{2(T-1)} - \frac{1}{2s(T-1)} + \frac{(s-K)^2}{s(T-1)} = 0 \ .$$

Combining these two observations with (10) and (11), we obtain

$$d_{t+\tau-1}\left(\Delta_1^*\right) + d_{t+\tau}\left(\Delta_2^*\right) \leq 2b - \frac{1}{T-1} - \ell^o(b,\tau,K) + \ell^e\left(b,s-\tau+1,K'\right) + \frac{2t+\tau+K}{T-1} \ .$$

Hence,

$$\begin{aligned}\Upsilon(b) &\equiv 2\left(b + \frac{t+\tau-1}{T-1}\right) - \left[d_{t+\tau-1}\left(\Delta_1^*,b\right) + d_{t+\tau}\left(\Delta_2^*,b\right)\right] \\ &\geq \frac{\tau-K-1}{T-1} + \ell^o(b,\tau,K) - \ell^e\left(b,s-\tau+1,K'\right) \ .\end{aligned}$$

Tedious calculations (available in the online Appendix) reveal that the right-hand side of the above inequality is nonnegative when (12) is true. This implies that $\Upsilon(b) \geq 0$, contrary to (9). As $t \leq T-s$ was chosen arbitrarily, this proves that the chain cannot break when the agent is offered an element of $\mathcal{D}^*_{s+1,t}(b)$, for every $t \leq T-s$.

Proof of Lemma 4

Let $\Delta = (\delta_t, \ldots, \delta_{t+K}) \in \mathcal{D}^*_{s,t}(b)$ satisfy the chain property. First of all, in order to simplify notation, we normalize t to 0. Thus, $\Delta = \{\delta_0, \ldots, \delta_K\}$. Moreover, in what follows we will indulge in a slight abuse of notation and define T_k and t_k as follows:

$$\begin{aligned} T_k &\equiv \left\{\tau \in \mathbf{T}_{s,t} : d_\tau\left(\Delta,b\right) = \delta_k\right\} \ , \\ t_k &\equiv \max\left\{\tau : \tau \in T_k\right\} \ . \end{aligned}$$

for each $k = 0, \ldots, K$.

Any delegation set is trivially top loaded if $s \leq 2$; so suppose that $s > 2$.

Suppose that

$$\delta_{k+1} < b + \frac{t_k+1}{T-1} \tag{13}$$

for every $k < K$. If Δ is not top loaded then there is $k < K$ such that $|T_k| > 1$. By (13), this implies that $\delta_k < b + \frac{t_k-1}{T-1}$. As Δ satisfies the chain property, this in turn implies that

$$\delta_{k+1} = 2b + \frac{2t_k}{T-1} - \delta_k > b + \frac{t_k+1}{T-1}:$$

contrary to (13).

What remains to be proved, therefore, is that (13) is true. We do so with a series of claims.

Claim 1: $\delta_{k+1} \leq b + \frac{t_k+1}{T-1}, \forall k < K$.

Let $k < K$ and suppose, contrary to Claim 1, that $\delta_{k+1} > b + \frac{t_k+1}{T-1}$. This implies that there is an integer $q \geq 1$ such that

$$b + \frac{t_k + q}{T-1} < \delta_{k+1} \leq b + \frac{t_k + q + 1}{T-1} \ .$$

Combining this with the chain property:

$$\delta_k = 2\left(b + \frac{t_k}{T-1}\right) - \delta_{k+1} \ ,$$

we obtain

$$b + \frac{t_k - q - 1}{T-1} \leq \delta_k < b + \frac{t_k - q}{T-1} \ .$$

We need to consider two cases separately:

• *Case 1: Either $k \neq K-1$ or $t_{K-1} + q \leq s-1$.* Let $\upsilon > 0$ be such that the agent is indifferent between δ_{k+1} and $\delta_{k+1} - \upsilon$ in state $t_k + q$. Note that, by construction, this implies that

$$\delta_k = b + \frac{t_k - q}{T-1} - \frac{\upsilon}{2} \ , \tag{14}$$

and that the agent is indifferent between δ_k and $\delta_k + \upsilon$ in state $t_k - q$.

Suppose now that the principal deviates from Δ to $\Delta \cup \{\delta_k + \upsilon, \delta_{k+1} - \upsilon\}$. The agent takes the same decisions in states $\{t : t \leq t_k - q \text{ or } t \geq t_k + q + 1\}$ as before the principal's deviation; now takes decision $\delta_k + \upsilon$ in states $\{t : t_k - q + 1 \leq t \leq t_k\}$; and now takes decision $\delta_{k+1} - \upsilon$ in states $\{t : t_k + 1 \leq t \leq t_k + q\}$.

We now decompose the change in the principal's expected loss after his deviation. First, the change in loss in states t_k and $t_k + q$ is given by

$$\left(\delta_k + \upsilon - \frac{t_k}{T-1}\right)^2 - \left(\delta_k - \frac{t_k}{T-1}\right)^2 + \left(\delta_{k+1} - \upsilon - \frac{t_k + q}{T-1}\right)^2 - \left(\delta_{k+1} - \frac{t_k + q}{T-1}\right)^2$$

$$= 2\upsilon\left(\upsilon + \delta_k - \delta_{k+1} + \frac{q}{T-1}\right) = 2\upsilon\left[2\left(\delta_k - b - \frac{t_k}{T-1} + \frac{\upsilon}{2}\right) + \frac{q}{T-1}\right] = -\frac{2\upsilon q}{T-1} < 0 \ ,$$

where the third equality follows from (14).

The change in the principal's loss in the other states is given by

$$\sum_{i=1}^{q-1}\left\{\left(\delta_k + \upsilon - \frac{t_k - i}{T-1}\right)^2 - \left(\delta_k - \frac{t_k - i}{T-1}\right)^2 + \left(\delta_{k+1} - \upsilon - \frac{t_k + i}{T-1}\right)^2 - \left(\delta_{k+1} - \frac{t_k + i}{T-1}\right)^2\right\} \ .$$

For each $i = 1, \ldots, q-1$, the bracketed term is equal to

$$2\upsilon\left(\upsilon + \delta_k - \delta_{k+1} + \frac{2i}{T-1}\right) = 2\upsilon\left[2\left(\delta_k - b - \frac{t_k}{T-1} + \frac{\upsilon}{2}\right) + \frac{2i}{T-1}\right] = -\frac{4\upsilon(q-i)}{T-1} < 0 \ .$$

As a consequence, the deviation from Δ is strictly profitable to the principal: a contradiction to Δ being an ODS.

• *Case 2: $k = K-1$ and $t_{K-1}+q > s-1$.* Define $\tilde{v}$ such that the agent is indifferent between δ_K and $\delta_K - \tilde{v}$ in state $s-1$. Let $\tilde{q} \equiv s - t_{K-1} - 1$. By construction, the agent is indifferent between δ_k and $\delta_k + \tilde{v}$ in state $t_{K-1} - \tilde{q}$. We can then repeat the same argument as in Case 1 — just substitute $\tilde{q}$ and $\tilde{v}$ for q and v, respectively — to obtain the same contradiction.

Claim 2: If $\delta_{k+1} \leq b + \frac{t_k+1}{T-1}, \forall k < K$ then $\delta_{k+1} < b + \frac{t_k+1}{T-1}, \forall k < K$.

Suppose that there is k such that

$$\delta_{k+1} = b + \frac{t_k + 1}{T-1} \ . \tag{15}$$

To obtain a contradiction, we will first prove that (15) implies that

$$\delta_l = b + \frac{t_l - 1}{T-1} \tag{16}$$

for every $0 \leq l < K$. The chain property implies that

$$\delta_k = 2b + \frac{2t_k}{T-1} - \delta_{k+1} = b + \frac{t_k - 1}{T-1} \ .$$

From Claim 1, this in turn implies that $t_k - 1$ is the smallest state in which the agent takes δ_k. Indeed, if she took decision δ_k in state $t < t_k - 1$ then

$$\delta_k = b + \frac{t_k - 1}{T-1} > b + \frac{t}{T-1} \geq b + \frac{t_{k-1} + 1}{T-1} \ ,$$

contrary to Claim 1; and if $t_k = t_{k-1} + 1$ then $\delta_{k-1} = b + \frac{t_{k-1}}{T-1}$, contrary to the chain property.

Consequently, $t_{k-1} = t_k - 2$ and the agent is indifferent between δ_{k-1} and δ_k in state $t_k - 2$. This in turn implies that

$$\delta_{k-1} = 2b + \frac{2t_{k-1}}{T-1} - \delta_k = b + \frac{t_{k-1} - 1}{T-1} \leq b + \frac{t_{k-2} + 1}{T-1} \ ,$$

where the last inequality follows from Claim 1.

We can then proceed recursively to obtain (16) for $0 \leq l \leq k$. Furthermore,

$$\delta_{k+1} = b + \frac{t_k + 1}{T-1} = 2b + \frac{2t_{k+1}}{T-1} - \delta_{k+2}$$

implies that

$$\delta_{k+2} = b + \frac{2t_{k+1} - t_k - 1}{T-1} \leq b + \frac{t_{k+1} + 1}{T-1} \ ,$$

so that $t_{k+1} \leq t_k + 2$. But we must have $t_{k+1} = t_k + 2$ for $k < K-1$: for if not, then $t_{k+1} = t_k + 1$. This would imply that the agent takes her ideal decision, $b + \frac{t_k+1}{T-1} = \delta_{k+1}$

in state t_{k+1}, and could therefore not be indifferent between δ_{k+1} and δ_{k+2}, contrary to the chain property. Thus,

$$\delta_{k+2} = b + \frac{t_{k+1}+1}{T-1} = 2b + \frac{2t_{k+2}}{T-1} - \delta_{k+3} \ .$$

We can then proceed recursively to obtain (16) for $k \leq l \leq K$.

Combined with the chain property, (15) also implies that $|T_k| = 2$ for $1 \leq k \leq K-1$, and $|T_k| \leq 2$ for $k = 0, K$. We then distinguish between two cases:

• *Case 1:* $|T_0| = 1$. In this case, the agent takes $\delta_0 = b - \frac{1}{T-1}$ in state 0; takes decision $\delta_i = b + \frac{2i-1}{T-1}$ in states $2i-1$ and $2i$: for every $0 < i < K$; and takes $\delta_K = b + \frac{2K-1}{T-1}$ at every state in T_K.

Note, also, that $b \geq \frac{1}{T-1}$ — else $\delta_0 < 0$, and the principal could improve on Δ by raising δ_0 — and that K must be odd.

If $|T_K| = 1$ then $s = 2K$ and the expected loss incurred by the principal is

$$\lambda(\Delta, b) \equiv K\left[b^2 + \left(b - \frac{1}{T-1}\right)^2\right] \ .$$

Consider the delegation set $\Delta^* \equiv \left\{\delta_0^*, \ldots, \delta_{s-1}^*\right\}$, where

$$\delta_t^* \equiv b + \frac{2t-1}{2(T-1)} : \forall t = 0, \ldots, s-1 \ .$$

It is easy to check that the loss incurred by the principal when he chooses Δ^* is

$$\lambda(\Delta^*, b) \equiv s\left[b - \frac{1}{2(T-1)}\right]^2 = \lambda(\Delta, b) - \frac{s}{4(T-1)^2} < \lambda(\Delta, b) \ ,$$

contrary to the supposition that Δ is an ODS.

If $|T_K| = 2$ then suppose that the principal deviates to $\Delta'_\varepsilon = \left\{\delta'_0, \ldots, \delta'_K\right\}$, where $\delta'_i = \delta_i + \varepsilon$ if i is even, $\delta'_i = \delta_i - \varepsilon$ if i is odd, and $\varepsilon > 0$. The change in the principal's loss is then $\varepsilon\left[(2K+1)\varepsilon - 2b\right]$, which is negative for every small enough ε: a contradiction.

• *Case 2:* $|T_0| = 2$. In this case, the agent takes decision $\delta_i = b + \frac{2i}{T-1}$ in states $2i$ and $2i+1$: for every $i = 0, \ldots, K-1$; and she takes decision $\delta_K = b + \frac{2K}{T-1}$ at states in T_K. Moreover, K must be even.

If $|T_K| = 1$ then the change in the principal's loss if he deviates to Δ'_ε (as defined above) is $\varepsilon\left[(2K+1)\varepsilon - 2b\right]$, which is negative if ε is small enough.

If $|T_K| = 2$ then $s = 2(K+1)$ and the principal's loss is

$$\lambda(\Delta, b) = (K+1)\left[b^2 + \left(b - \frac{1}{T-1}\right)^2\right] .$$

It is easy to check that a principal who offers Δ^* (as defined above) loses

$$\lambda(\Delta^*, b) \equiv s\left[b - \frac{1}{2(T-1)}\right]^2 = \lambda(\Delta, b) - \frac{s}{4(T-1)^2} < \lambda(\Delta, b) \ :$$

a contradiction. This completes the proof of Claim 2.

Combining Claims 1 and 2, we obtain (13).

Proof of Theorem 1

If $b \leq \frac{1}{2(T-1)}$, the result is obvious. Suppose $b > \frac{1}{2(T-1)}$. Part (i) is a direct consequence of ($\mathbf{H_2}$) and Lemmas 2-4. Note, however, that more is needed to establish parts (ii) and (iii) because Lemma 2 only establishes necessity.

We know from Lemma 1 that there is an ODS for every $b > \frac{1}{2(T-1)}$. Let Δ be an ODS. Define $\Delta^*(b, K)$ as the unique solution to the principal's minimization problem in (7): that is, $\Delta^*(b, K)$ is optimal among the delegation sets of cardinality $(K+1)$ satisfying the chain property and top loading.

Suppose, first, that $b \in \left(b^{\min}(K,T), b^{\max}(K,T)\right)$, but that Δ contains $K'+1$ decisions where $K' \neq K$. From Lemma 2, this implies that $b \in \left[b^{\min}(K',T), b^{\max}(K',T)\right]$: a contradiction. Consequently, Δ must be equal to $\Delta^*(b, K)$.

Now suppose that $b = b^{\min}(K,T)$. Applying Lemma 2, an ODS must contain either K or $K+1$ decisions. The computations in the proof of Lemma 2 reveal that, when $b = b^{\min}(K,T)$, $\Delta^*(b,K)$ and $\Delta^*(b, K-1)$ yield the same expected loss to the principal. Therefore, $\mathcal{D}_T^*(b)$ consists of $\Delta^*(b, K-1)$ and $\Delta^*(b,K)$. A parallel argument establishes the result when $b = b^{\max}(K,T)$.

APPENDIX B: Proofs of Propositions, Corollaries and Observations

Proof of Corollary 1

Let $\Delta \in \mathcal{D}_T^*(b)$, and let $|\Delta| - 1 = K$. Lemma 2 implies that $b \in \left[b^{\min}(K,T), b^{\max}(K,T)\right]$.

(i) Necessity is easy to establish: If $b \leq \frac{1}{2(T-1)}$ then the first-best decision rule is incentive compatible, so $\Delta \in \mathcal{D}_T^*(b)$ implies that

$$d_t(\Delta, b) = \frac{t}{T-1} \ , \forall t \in \mathbf{T} \ .$$

To establish sufficiency, suppose that $b > \frac{1}{2(T-1)}$. Lemma 2 implies that we have to show that $\ell(b,T,K) > 0$ if t is even, and $2b - \ell(b,T,K) > \frac{1}{T-1}$ if t is odd. We must distinguish between four cases:

• Case 1: *t even and K even.* Since ℓ^e is strictly increasing in b ($K < T$),

$$\ell^e(b,T,K) > \ell^e\left(b^{\min}(K,T), T, K\right) = \frac{T-K-1}{2(T-1)}$$

for all $b > b^{\min}(K,T)$. If $K < T-1$, we directly obtain the result from the equation equation because $b \geq b^{\min}(K,T)$. If $K = T-1$ then $b^{\min}(K,T) = \frac{1}{2(T-1)}$; so $b > b^{\min}(K,T)$. Thus, $\ell^e(b,T,K) > \ell^e\left(b^{\min}(K,T), T, K\right) = 0$.

• Case 2: *t odd and K even.* Since $2b - \ell^e(b,T,K)$ is strictly increasing in b,

$$2b - \ell^e(b,T,K) > 2b - \ell^e\left(b^{\min}(K,T),T,K\right) = \frac{T-K-1}{2(T-1)} + \frac{1}{T-1}$$

for all $b > b^{\min}(K,T)$. The argument is then the same as in Case 1.

• Case 3: *t even and K odd.* Since ℓ^o is strictly increasing in b,

$$\ell^o(b,T,K) > \ell^o\left(b^{\min}(K,T),T,K\right) = \frac{T-K-1}{2T}$$

for all $b > b^{\min}(K,T)$. The argument is then the same as in Case 1.

• Case 4: *t odd and K odd.* Since $2b - \ell^o(b,T,K)$ is strictly increasing in b,

$$2b - \ell^o(b,T,K) > 2b - \ell^o\left(b^{\min}(K,T),T,K\right) = \frac{(T-K-1)(T+1)}{2(T-1)} + \frac{1}{T-1}$$

for all $b > b^{\min}(K,T)$. The argument is then the same as in Case 1.

(ii) Let $\Delta = \{\delta_0, \ldots, \delta_K\}$. For every $t \in \mathbf{T}$, there is k such that $t \in T_{k+1}(\Delta, b)$. Claims 1 and 2 in the proof of Lemma 4 then imply that

$$d_t(\Delta,b) = \delta_{k+1} < b + \frac{t_k(\Delta,b)+1}{T-1} \leq b + \frac{t}{T-1} \ .$$

(iii) This follows immediately from the requirement that the agent be indifferent between δ_0 and δ_1 in state 0.

(iv) The condition in the premise implies that the ODS contains T decisions, the largest of which is 1 if $b = 1/2(T-1)$, and must offer a larger decision when the bias is larger, but small enough that the ODS contains T decisions. The agent must be indifferent between the two highest decisions in state $\frac{T-2}{T-1}$ for bias in this range; so $d_{T-2}(\Delta,b) < b + \frac{T-2}{T-1} < 1$.

Now suppose that $b > 1/(T-1)$, so $K < T-1$. The second highest decision in the ODS is then bounded above by the limit of δ_{K-1} as b approaches $\frac{T-K+1}{2(T-1)}$. We can show that the principal expects to lose less if the agent takes decision δ_{K-1} than if she takes any decision above 1, conditional on the state being at least $K/(T-1)$. Consequently, the maximal decision in an ODS must be less than 1 whenever $b > 1/(T-1)$.

Proof of Observation 1

We distinguish between two cases:

(a) $b > 1/2$. In this case, we have $b > \frac{1}{2} = \lim_{T\to\infty} \frac{T}{2(T-1)}$. Consequently, $b > T/2(T-1)$ for large enough T. Theorem 1 then implies that there is a unique ODS, namely $\{1/2\}$, as in Holmstrom's model.

(b) $b \in (0, 1/2]$. As $T/2(T-1)$ is a strictly decreasing function that converges to $1/2$ as $T \to \infty$, we have

$$b \leq \frac{1}{2} < \frac{T}{2(T-1)} \ , \forall T \in \mathbb{N} \ . \tag{17}$$

For every $T \in \mathbb{N}$, let

$$\widetilde{K}(T) \equiv \left\{ K \in \mathbb{N} : K \leq T-1 \text{ and } \frac{T-K}{2(T-1)} \leq b \leq \frac{T-K+1}{2(T-1)} \right\} \ .$$

Lemma 2, Theorem 1, and (17) imply that, for every T and every $\Delta_T^*(b) \in \mathcal{D}_T^*(b)$, there exists $K_T \in \widetilde{K}(T)$ such that $\Delta_T^*(b)$ is of the form $\Delta_T^*(b) = \{\delta_0, \ldots, \delta_{K_T}\}$, and

$$\delta_\tau = \begin{cases} \ell(b, T, K_T) + \frac{\tau}{T-1} & \text{if } \tau \text{ is even,} \\ 2b - \ell(b, T, K_T) + \frac{\tau - 1}{T-1} & \text{if } \tau \text{ is odd,} \end{cases}$$

where ℓ is defined before Lemma 2 in Section 2.3.

We want to prove that each $\Delta_T^*(b) \in \mathcal{D}_T^*(b)$ becomes arbitrarily close to $\Delta_\infty^*(b)$ as $T \to \infty$. To do so, we will decompose $\mathcal{D}_T^*(b)$ as follows: $\mathcal{D}_T^*(b) = \mathcal{D}_T^e(b) \cup \mathcal{D}_T^o(b)$, where $\mathcal{D}_T^e(b)$ [resp. $\mathcal{D}_T^o(b)$] is the class of ODSs $\Delta_T^*(b) = \{\delta_0, \ldots, \delta_{K_T}\}$ such that K_T is even [resp. odd].

Suppose, first, that $\Delta_T^*(b) \in \mathcal{D}_T^e(b)$ (i.e., K_T is even). Let

$$E^T \equiv \left\{ 0, \frac{2}{T-1}, \frac{4}{T-1}, \ldots, \frac{K_T - 2}{T-1}, \frac{K_T}{T-1} \right\} \ ,$$
$$O^T \equiv \left\{ \frac{1}{T-1}, \frac{3}{T-1}, \ldots, \frac{K_T - 3}{T-1}, \frac{K_T - 1}{T-1} \right\} \ ,$$

and note that we can express $\Delta_T^*(b)$ as

$$\Delta_T^*(b) = \left(\{\ell^e(b, T, K_T)\} + E^T \right) \bigcup \left(\left\{ 2b - \ell^e(b, T, K_T) - \frac{1}{T-1} \right\} + O^T \right) \ . \tag{18}$$

Since $K_T \in \widetilde{K}(T)$, we have

$$\frac{T - K_T}{2(T-1)} \leq b \leq \frac{T - K_T + 1}{2(T-1)}$$

and therefore

$$T - 2(T-1)b \leq K_T \leq T - 2(T-1)b + 1 \ . \tag{19}$$

This in turn implies that

$$\frac{T - 2(T-1)b}{T-1} \leq \frac{K_T}{T-1} \leq \frac{T - 2(T-1)b + 1}{T-1} \ .$$

Since both the left-hand and right-hand sides of this inequality converge to $1 - 2b$, we obtain that $K_T/(T-1)$ and $(K_T - 1)/(T-1)$ converge to $1 - 2b$ as $T \to \infty$. This in turn implies that both E^T and O^T converge to the interval $[0, 1-2b]$ as $T \to \infty$.

The definition of ℓ^e and (19) imply that

$$\frac{T-2(T-1)b}{T}b \leq \frac{K}{T}b \leq \frac{T-2(T-1)b+1}{T}b \ ,$$

where the left-hand and right-hand sides converge to $(1-2b)b$ as T becomes arbitrarily large. This proves that $\frac{K_T}{T}b$ also converges to $(1-2b)b$ as $T \to \infty$. Using (19) again, we obtain

$$\frac{[2(T-1)b-1]^2}{2T(T-1)} \leq \frac{(T-K_T)^2}{2T(T-1)} \leq \frac{[2(T-1)b]^2}{2T(T-1)}$$

where the left-hand and right-hand side both approach $4b^2/2 = 2b^2$. This proves that $\frac{(T-K_T)^2}{2T(T-1)}$ converges to $2b^2$ as $T \to \infty$.

Thus, $\ell^e(b, T, K_T)$, and therefore $2b - \ell^e(b, T, K_T)$, converge to b as $T \to \infty$. By (18), the limit of $\Delta^*_T(b)$ can then be expressed as

$$(\{b\} + [0, 1-2b]) \cup (\{b\} + [0, 1-2b]) \ .$$

We have therefore proved that every $\Delta^*_T(b) \in \mathcal{D}^e_T(b)$ converges to $[b, 1-b]$ — which is equal to $\Delta^*_\infty(b)$ — as $T \to \infty$.

A parallel argument shows that every $\Delta^*_T(b) \in \mathcal{D}^o_T(b)$ converges to $\Delta^*_\infty(b)$ as $T \to \infty$. This proves that $\lim_{T\to\infty} \Delta^*_T(b) = \Delta^*_\infty(b)$ for all $\Delta^*_T(b) \in \mathcal{D}^*_T(b)$.

Proof of Proposition 1

We start with small changes. Theorem 1 implies that the agent is indifferent between δ_t and δ_{t+1} in state $t < K$. If the initial bias satisfies $b_0 \in \left(\frac{T-K_0}{2(T-1)}, \frac{T-K_0+1}{2(T-1)}\right)$ then a small enough increase in b leaves K fixed at K_0. However, the agent now strictly prefers δ_{t+1} over δ_t; so the principal must increase the t'th and/or the next highest decision in the ODS to restore the agent's indifference in state t. Theorem 1 implies that the ODS varies continuously for increases in bias in this range. In sum, if $K_0 > 0$ (so delegation is valuable) then a small enough increase in b from a generic starting point shifts the ODS to the right.

Now consider the principal's loss at a critical value of b, say $b^{\max}(K, T)$. The chain property implies that the principal then makes the same loss in states $t \geq K$ if the agent takes δ_{K-1} or δ_K; so $\{\delta_0, \ldots, \delta_{K-1}\}$ is also an ODS at this critical bias. Thus, a marginal increase in b, starting from $b_0 \in \left\{\frac{t}{2(T-1)}\right\}_{t>1}$, induces the principal to drop δ_K from the ODS, and to change the smaller decisions continuously. In particular, if B is a critical bias then

$$\lim_{b \searrow B} \Delta(b) \subset \lim_{b \nearrow B} \Delta(b).$$

This property mimics the reduction in discretion in the continuous-state model.

Consider an increase in bias from b_0 to b_1 which is large enough that $\Delta(b_1)$ contains fewer decisions than $\Delta(b_0)$. Although $\Delta(b_0)$ and $\Delta(b_1)$ are not ordered by discretion, the principal offers more decisions to a less biased agent. We will now argue that a large enough increase in b reduces the maximal decision and raises the minimal decision in the ODS.

We start with the first claim. It is sufficient to show that an increase in b which reduces the number of decisions from $K_0 + 1$ to K_0 reduces the maximal decision.

If K_0 is even then the requisite condition is equivalent to

$$\ell^e(b_0, T, K_0) + \frac{K_0}{T-1} > 2b_1 - \ell^o(b_1, T, K_0 - 1) + \frac{K_0 - 2}{T-1}.$$

Substituting out for ℓ^e and ℓ^o, this condition is equivalent to $b_1 - b_0 < \frac{T+K_0}{2(T-1)}$. Now

$$\frac{T - K_0}{2(T-1)} < b_0 < b_1 < \frac{T - K_0 + 2}{2(T-1)},$$

which implies that the condition holds.

If K_0 is odd and $K_0 > 1$ then the requisite condition is equivalent to

$$2b_0 - \ell^o(b_0, T, K_0) > \ell^e(b_0, T, K_0 - 1)$$

which, on rearrangement, yields $b_1 - b_0 < 1/(T-1)$. The bounds on b_0 and b_1 above imply that this condition holds. If $K_0 = 1$ then Theorem 1 directly implies that $\delta_1 > 1/2$. These observations imply our claim that the maximal decision in $\Delta(b_1)$ exceeds the maximal decision in $\Delta(b_0)$ whenever $b_1 - b_0$ is large enough that the principal reduces the number of decisions in the ODS.

We will now argue that the minimal decision in $\Delta(b_1)$ exceeds the minimal decision in $\Delta(b_0)$ by considering an increase in b which reduces the number of decisions from $K_0 + 1$ to K_0. The argument requires us to distinguish between cases in which K_0 is odd and even. In the latter case, Lemma 2 implies that the minimal decision in $\Delta(b_1)$ exceeds the minimal decision in $\Delta(b_0)$ if and only if

$$K_0 b_0 - \frac{T}{2(T-1)} + \frac{(T - K_0)^2}{2(T-1)} < (2T - K_0)\, b_1 - \frac{1}{2} - \frac{(T - K_0 + 1)^2}{2(T-1)}.$$

Now

$$K_0 b_0 \leq \frac{K_0(T - K_0 + 1)}{2(T-1)} \text{ and } (2T - K_0) b_1 \geq \frac{(2T - K_0)(T - K_0 + 1)}{2(T-1)}.$$

Substituting above yields the requisite inequality.

If K_0 is odd then Lemma 2 implies that the minimal decision in $\Delta(b_1)$ exceeds the minimal decision in $\Delta(b_0)$ if and only if

$$(2T - K_0 - 1)b_0 - (K_0 - 1)b_1 < \frac{(T - K_0)^2}{T-1} + \frac{T - K_0}{T-1} + \frac{1}{T-1}.$$

Now

$$(2T - K_0 - 1)b_0 \leq \frac{(2T - K_0 - 1)(T - K_0 + 1)}{2(T-1)} \text{ and } (K_0 - 1)b_1 \geq \frac{(K_0 - 1)(T - K_0 + 1)}{2(T-1)}.$$

Substituting above yields the requisite inequality.

This pair of observations implies our claim that the minimal decision in $\Delta(b_1)$ exceeds the minimal decision in $\Delta(b_0)$.

Proof of Proposition 3

(i) Replacing the expert with an amateur must make the agent's decisions less state-sensitive; so the expert must outperform any amateur if the agent would take first-best decisions in every state: that is, when $b \leq \frac{1}{2(T-1)}$.

If $b > \frac{T}{2(T-1)}$ then the principal loses $\frac{T^2+T}{12(T-1)^2}$ if he delegates to an expert. The principal can improve on appointing an expert if and only if he can offer some amateur a two-decision delegation set.

Suppose that the principal offers $\{\delta_0, \delta_1\}$ to an amateur, who takes δ_0 if and only if the state is in $E \equiv \{0, ..., \tau - 1\}$. The principal is then best off if the amateur cannot distinguish between the states in E, and is indifferent between δ_0 and δ_1: so

$$\delta_1 = 2b + \frac{\tau - 1}{T - 1} - \delta_0 \ .$$

For fixed δ_0 and τ, the principal expects to lose

$$\lambda\left(\delta_0, \tau, b\right) \equiv \frac{1}{T-1}\left[\sum_{t=0}^{\tau-1}(\delta_0 - \frac{t}{T-1})^2 + \sum_{t=\tau}^{T-1}\left(\delta_0 - 2b - \frac{\tau-1}{T-1} + \frac{t}{T-1}\right)^2\right].$$

For fixed τ, the principal minimizes his loss by offering the amateur

$$\delta_0^*(\tau, b) \equiv \frac{2(T-\tau)}{T}b - \frac{1}{2} + \frac{\tau - 1}{T-1} \text{ and } \delta_1^*(\tau, b) = \frac{2\tau}{T}b + \frac{1}{2} \ .$$

Now $\lambda\left[\delta_0^*(\tau, b), \tau, b\right]$ is increasing in b; so $\lambda\left[\delta_0^*(\tau, b), \tau; b\right] > \lambda\left[\delta_0^*(\tau, b), \tau, \frac{T}{2(T-1)}\right]$ in the relevant range. We will argue that the principal is indifferent between appointing the expert and an amateur when $b = T/2(T-1)$, which will then imply that the expert outperforms the amateur for larger bias. Substituting for b yields

$$\delta_0^*\left(\tau, \frac{T}{2(T-1)}\right) = \frac{1}{2} \text{ and } \delta_1^*\left(\tau, \frac{T}{2(T-1)}\right) = \frac{\tau}{T-1} + \frac{1}{2} \ .$$

The expert takes decision 1/2 in all states; so the amateur outperforms the expert when $b = T/2(T-1)$ if and only if the principal loses less in states $t \geq \tau$: that is, if there is τ such that

$$L(\tau) \equiv \sum_{t=\tau}^{T-1}\left(\frac{1}{2} + \frac{\tau}{T-1} - \frac{t}{T-1}\right)^2 - \sum_{t=\tau}^{T-1}\left(\frac{1}{2} - \frac{t}{T-1}\right)^2 < 0 \ .$$

Rearranging terms reveals that $L(\tau) = 0$ for every τ, proving this part.

(ii) Suppose, first, that $\frac{1}{T-1} < b < \frac{1}{2}$, and let $\Delta = \{\delta_0, \ldots, \delta_K\} \in \mathcal{D}_T^*(b)$. Theorem 1 implies that $K > 1$.

Let E denote the event $\{K-2, K-1\}$, and consider an amateur who can distinguish between every state except for those in E. Define δ_E as the decision which leaves this amateur indifferent with δ_K in event E: viz.

$$\delta_E \equiv 2b + \frac{2T-3}{T-1} - \delta_K \ .$$

For fixed T, we need to distinguish between cases in which K is odd and K is even.

• If K is even (so $K-1$ is odd) then, by Theorem 1,

$$\begin{aligned}
\delta_{K-2} &= \frac{K}{T}b + \frac{K^2+T^2-5T}{2T(T-1)} \ , , \\
\delta_{K-1} &= \frac{2T-K}{T}b - \frac{K^2-4KT+T^2+3T}{2T(T-1)} \\
\text{and } \delta_K &= \frac{K}{T}b + \frac{K^2+T^2-T}{2T(T-1)};
\end{aligned}$$

so

$$\delta_E = \frac{2T-K}{T}b + \frac{4KT-5T-K^2-T^2}{2T(T-1)} > \delta_{K-2} \ ,$$

where the inequality implies that the agent prefers δ_{K-3} over δ_E in state $K-3$.

If the principal appoints an expert and offers Δ then, in event E, he loses

$$\begin{aligned}
&\left(\delta_{K-2} - \frac{K-2}{T-1}\right)^2 + \left(\delta_{K-1} - \frac{K-1}{T-1}\right)^2 \\
&= \left[\frac{K}{T}b + \frac{(T-K)^2-T}{2T(T-1)}\right]^2 + \left[\frac{2T-K}{T}b - \frac{(T-K)^2+T}{2T(T-1)}\right]^2 \ .
\end{aligned}$$

Now suppose that the principal offers an amateur a delegation set identical to Δ, except that δ_{K-2} and δ_{K-1} are replaced by δ_E. By construction, the amateur takes the same decision as the expert except in event E, where she takes δ_E. The principal's loss in event E is then

$$\left(\delta_E - \frac{K-2}{T-1}\right)^2 + \left(\delta_E - \frac{K-1}{T-1}\right)^2 = \left[\frac{2T-K}{T}b - \frac{(T-K)^2+T}{2T(T-1)}\right]^2 + \left[\frac{2T-K}{T}b - \frac{(T-K)^2+3T}{2T(T-1)}\right]^2 \ .$$

The principal prefers to appoint this amateur and offer her δ_E if and only if

$$\left[\frac{2T-K}{T}b - \frac{(T-K)^2+3T}{2T(T-1)}\right]^2 < \left[\frac{K}{T}b + \frac{(T-K)^2-T}{2T(T-1)}\right]^2 \ . \tag{20}$$

Now, Theorem 1 implies that $b < \frac{T-K+1}{2(T-1)} < \frac{(T-K)^2+T}{2(T-1)(T-K)}$; so

$$\frac{2T-K}{T}b - \frac{(T-K)^2+3T}{2T(T-1)} < \frac{K}{T}b + \frac{(T-K)^2-T}{2T(T-1)} \ . \tag{21}$$

Furthermore,

$$\delta_{K-2} - \frac{K-2}{T-1} = \frac{K}{T}b + \frac{(T-K)^2-T}{2T(T-1)} > 0$$

because $\delta_{K-2} = d_{K-2}(\Delta, b) > \frac{K-2}{T-1}$ (by Corollary 1).

There are two cases to consider. If the left-hand side of (20) is positive then (21) implies that the principal prefers to appoint this amateur and offer her δ_E; and if the left-hand side of (20) is negative then

$$\frac{(T-K)^2+3T}{2T(T-1)} - \frac{2T-K}{T}b < \frac{K}{T}b + \frac{(T-K)^2-T}{2T(T-1)}$$

if and only if $b > 1/(T-1)$. We therefore conclude that the principal prefers to appoint this amateur and offer her δ_E if $b > 1/(T-1)$.

• If K is odd (so $K-1$ is even) then

$$\begin{aligned}
\delta_{K-2} &= \frac{K+1}{T}b + \frac{K^2+T^2-5T-1}{2T(T-1)} \ , \\
\delta_{K-1} &= \frac{2T-K-1}{T}b - \frac{K^2+T^2-4KT+3T-1}{2T(T-1)} \ , \\
\delta_K &= \frac{K+1}{T}b + \frac{K^2+T^2-T-1}{2T(T-1)}, \\
\delta_E &= \frac{2T-K-1}{T}b + \frac{4KT-5T-K^2-T^2+1}{2T(T-1)} < \delta_{K-2} \ ,
\end{aligned}$$

where the last inequality exploits $b < \frac{T-K+1}{2(T-1)}$, and implies that the agent must take a decision $\delta \geq \delta_{K-2}$ in event E if she continues to take δ_t in states $t < K-2$.

If $\delta = \delta_{K-2}$ then the amateur outperforms the expert if and only if the principal prefers the agent to take δ_{K-2} rather than δ_{K-1} in state $K-1$: that is, when $b > 1/(T-1)$. Now $\frac{1}{T-1} < \frac{T-K}{2(T-1)} = b^{\min}(K,T)$ if and only if $K < T-1$: so the amateur outperforms the expert for every $b \in \left(\frac{1}{T-1}, \frac{1}{2}\right)$.

Now consider cases in which $b \in \left(\frac{1}{2}, \frac{T}{2(T-1)}\right)$: so $K = 1$ (by Theorem 1). The expert would take $\frac{2(T-1)}{T}b - \frac{1}{2}$ in state 0, and $\frac{2}{T}b + \frac{1}{2}$ otherwise; so the principal expects to lose

$$\lambda(\Delta, b) = \frac{1}{T}\left\{\left[\frac{2(T-1)}{T}b - \frac{1}{2}\right]^2 + \left[\frac{2}{T}b + \frac{T-3}{2(T-1)}\right]^2 + \sum_{t=2}^{T-1}\left[\frac{t}{T-1} - \frac{2}{T}b - \frac{1}{2}\right]^2\right\} \ .$$

Consider an amateur who can't distinguish between states 0 and 1. A principal who offers this amateur the delegation set

$$\left\{\frac{2(T-2)}{T}b - \frac{T-3}{2(T-1)}, \frac{4}{T}b + \frac{1}{2}\right\}$$

expects to lose

$$\lambda_E \equiv \frac{1}{T}\left\{\left[\frac{2(T-2)}{T}b - \frac{1}{2}\right]^2 + \left[\frac{2(T-2)}{T}b - \frac{T-3}{2(T-1)}\right]^2 + \sum_{t=2}^{T-1}\left[\frac{t}{T-1} - \frac{4}{T}b - \frac{1}{2}\right]^2\right\} \ .$$

Subtracting terms yields

$$\lambda_E - \lambda(\Delta, b) = \frac{4(T-3)}{T^2} b \left[b - \frac{T}{2(T-1)} \right] .$$

Consequently, this amateur strictly outperforms the expert if and only if $T > 3$.

Proof of Proposition 4

Proof We denote the amateur who cannot distinguish between states 0 and 1 as 01. We start by arguing that the ODS for agent 01 cannot contain more than two decisions:

Suppose otherwise, and let the agent take δ_{K-1} in states $\tau = t+1, \ldots, u$, and δ_K in states $\tau = u+1, \ldots, T-1$, where $u > 1$. If the principal deviated to offering $\{\delta_0, ..., \delta_{K-1}\}$ then the agent would take the same decisions in states $\tau \leq u$; so T times the difference between the principal's loss with the new DS and the loss with the putative ODS equals

$$\begin{aligned}
& (\delta_{K-1} - \delta_K) \left[(T-1-u)(\delta_{K-1} + \delta_K) - \frac{2}{T-1} \sum_{\tau=u+1}^{T-1} \tau \right] \\
\leq \; & (\delta_{K-1} - \delta_K) \left[(T-1-u)(2b + \frac{2u}{T-1}) - T + \frac{u(u+1)}{T-1} \right] \\
< \; & (\delta_{K-1} - \delta_K) (u-1) \left(1 - \frac{u}{T-1} \right) :
\end{aligned}$$

where the weak inequality follows from $\delta_{K-1} < \delta_K$ and the supposition that the agent takes δ_{K-1} in state u, and the strict inequality follows from $b > 1/2$. Consequently, the ODS for 01 cannot contain more than two decisions.

If the ODS for 01 failed the chain property then

$$\delta_0 = \frac{t}{2(T-1)} < \frac{T+t-1}{2(T-1)} = \delta_1$$

for some state $t \geq 1$. However, the agent would then take δ_1 in state t because $b > 1/2$ implies that

$$\frac{T+t-1}{2(T-1)} < \frac{T+2t-1}{2(T-1)} < b + \frac{t}{T-1}.$$

Consequently, the ODS for 01 satisfies the chain property.

We now argue that the ODS for 01 must induce the agent to take δ_1 in every state $t > 1$. To see this suppose, per contra, that the ODS for 01 (say, Δ) induces the agent to take δ_0 in every state $\tau < t$ for some $t > 1$. The expert and 01 would take the same decision in each state if they were each offered Δ. Theorem 1 implies that the principal could improve on Δ by offering Δ^* to the expert, who would therefore outperform 01, contrary to Proposition 3(ii).

These arguments imply that the ODS for 01 is

$$\left\{ \frac{2(T-2)}{T} b - \frac{T-3}{2(T-1)}, \frac{4}{T} b + \frac{1}{2} \right\};$$

so the principal loses

$$\lambda^{01} = \frac{1}{T}\left[\frac{8(T-2)}{T}b^2 - \frac{4(T-2)}{T-1}b - \frac{T}{4} + C_2\right]$$
$$\text{where } C_2 \equiv \frac{1}{(T-1)^2}\sum_{t=0}^{T-1} t^2.$$

The corresponding loss for the agent is

$$\begin{aligned}\Lambda^{01} &= \frac{1}{T}\left[(\delta_0 - b)^2 + \left(\delta_0 - b - \frac{1}{T-1}\right)^2 + \sum_{t=2}^{T-1}\left(\delta_1 - b - \frac{t}{T-1}\right)^2\right] \\ &= \frac{1}{T}\left[2\delta_0\left(\delta_0 - 2b - \frac{1}{T-1}\right) + (T-2)\delta_1\left(\delta_1 - 2b - \frac{T+1}{T-1}\right) + C_2^A\right] \\ &= \frac{1}{T}\left[C_2^A - \frac{8(T-2)}{T}b^2 - Tb - \frac{T}{4}\right] \\ \text{where } C_2^A &\equiv \sum_{t=0}^{T-1}\left(b + \frac{t}{T-1}\right)^2.\end{aligned}$$

Theorem 1 implies that the principal loses

$$\lambda^* = \frac{1}{T}\left[\frac{4(T-1)}{T}b^2 - 2b - \frac{T}{4} + C_2\right]$$

if he appoints the expert. We can also compute the corresponding loss for the agent:

$$\Lambda^* = \frac{1}{T}\left[C_2^A - \frac{4(T-1)}{T}b^2 - Tb - \frac{T}{4}\right].$$

Comparing losses, we have

$$T\left(\lambda^* - \lambda^{01}\right) = -2(T-3)b\left(\frac{2}{T}b - \frac{1}{T-1}\right),$$

which is positive whenever $b < \frac{T}{2(T-1)}$; and

$$T\left(\Lambda^* - \Lambda^{01}\right) = \frac{4(T-3)}{T}b^2 > 0\ .$$

APPENDIX C: Examples

C.1 The three-state case

Suppose that the set of states is $\{0, \frac{1}{2}, 1\}$. The Table below describes the ODS (Δ^*) and the loss incurred by the principal (λ^*) and the expert (Λ^*) for generic bias:

b	Δ^*	λ^*	Λ^*
$b < \frac{1}{4}$	$0, \frac{1}{2}, 1$	0	b^2
$\frac{1}{4} < b < \frac{1}{2}$	$\frac{2}{3}b - \frac{1}{6}, \frac{4}{3}b + \frac{1}{6}, \frac{2}{3}b + \frac{5}{6}$	$\frac{8}{9}b^2 - \frac{4}{9}b + \frac{1}{18}$	$\frac{1}{3}b^2 + \frac{1}{18}$
$\frac{1}{2} < b < \frac{3}{4}$	$\frac{4}{3}b - \frac{1}{2}, \frac{2}{3}b + \frac{1}{2}$	$\frac{8}{9}b^2 - \frac{2}{3}b + \frac{1}{6}$	$\frac{1}{9}b^2 + \frac{1}{6}$
$b > \frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{6}$	$b^2 + \frac{1}{6}$

There are four possible amateurs: 01, 12, 02 and 012. Write Δ^E for the ODS for an amateur who cannot distinguish between states in some collection of events E; write λ^E and $\Lambda^E$for the losses incurred by the principal and agent respectively. We then have

b	Δ^{01}	λ^{01}	Λ^{01}
$b < \frac{3}{8}$	$\frac{1}{4}, 1$	$\frac{1}{24}$	$b^2 + \frac{1}{24}$
$\frac{3}{8} < b < \frac{3}{4}$	$\frac{2}{3}b, \frac{4}{3}b + \frac{1}{2}$	$\frac{8}{9}b^2 - \frac{2}{3}b + \frac{1}{6}$	$\frac{1}{9}b^2 + \frac{1}{6}$
$b > \frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{6}$	$b^2 + \frac{1}{6}$

and

b	Δ^{12}	λ^{12}	Λ^{12}
$b < \frac{3}{8}$	$0, \frac{3}{4}$	$\frac{1}{24}$	$b^2 + \frac{1}{24}$
$\frac{3}{8} < b < \frac{3}{4}$	$\frac{4}{3}b - \frac{1}{2}, \frac{2}{3}b + \frac{1}{2}$	$\frac{8}{9}b^2 - \frac{2}{3}b + \frac{1}{6}$	$\frac{1}{9}b^2 + \frac{1}{6}$
$b > \frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{6}$	$b^2 + \frac{1}{6}$

If $E = \{0, 2\}$ or $E = \{0, 1, 2\}$ then the agent must take the same decision in every state; so the agent takes $1/2$ in every state.

These calculations imply that the principal appoints an expert when delegation is valuable, confirming our claim in the discussion following Proposition 3. On the other hand, the agent strictly prefers to be an expert when $b < 1/2$, and is otherwise indifferent between being an expert and an amateur who cannot distinguish between states t and $t + 1$.

C.2 The four-state case

Suppose that the set of states is $\{0, \frac{1}{3}, \frac{2}{3}, 1\}$. In contrast to the three-state case, we only report the ODS and losses for agents mentioned in the text. We report the analogous statistics for other amateurs in Anesi and Seidmann (2011).

The Table below describes the ODS and the loss incurred by the principal and the expert for generic bias:

b	Δ^*	λ^*	Λ^*
$\frac{1}{6} < b < \frac{1}{3}$	$b - \frac{1}{6}, b + \frac{1}{6}$ $b + \frac{1}{2}, b + \frac{5}{6}$	$(b - \frac{1}{6})^2$	$\frac{1}{36}$
$\frac{1}{3} < b < \frac{1}{2}$	$\frac{1}{2}b, \frac{3}{2}b, \frac{1}{2}b + \frac{2}{3}$	$\frac{3}{4}b^2 - \frac{1}{3}b + \frac{1}{18}$	$\frac{1}{4}b^2 + \frac{1}{18}$
$\frac{1}{2} < b < \frac{2}{3}$	$\frac{3}{2}b - \frac{1}{2}, \frac{1}{2}b + \frac{1}{2}$	$\frac{3}{4}b^2 - \frac{1}{2}b + \frac{5}{36}$	$\frac{1}{4}b^2 + \frac{5}{36}$
$b > \frac{2}{3}$	$\frac{1}{2}$	$\frac{5}{36}$	$b^2 + \frac{5}{36}$

If $E = \{0, 1\}$ then

	Δ^{01}	λ^{01}	Λ^{01}
$b < \frac{1}{6}$	$\frac{1}{6}, \frac{2}{3}, 1$	$\frac{1}{72}$	$b^2 + \frac{1}{72}$
$\frac{1}{6} < b < \frac{1}{3}$	$\frac{1}{6}, b + \frac{1}{2}, b + \frac{5}{6}$	$\frac{1}{2}b^2 - \frac{1}{6}b + \frac{1}{36}$	$\frac{1}{2}b^2 + \frac{1}{36}$
$\frac{1}{3} < b < \frac{2}{3}$	$b - \frac{1}{6}, b + \frac{1}{2}$	$b^2 - \frac{2}{3}b + \frac{5}{36}$	$\frac{5}{36}$
$b > \frac{2}{3}$	$\frac{1}{2}$	$\frac{5}{36}$	$b^2 + \frac{5}{36}$

If $E = \{1, 2\}$ then

b	Δ^{12}	λ^{12}	Λ^{12}
$b < \frac{1}{4}$	$0, \frac{1}{2}, 1$	$\frac{1}{72}$	$b^2 + \frac{1}{72}$
$\frac{1}{4} < b < \frac{1}{2}$	$b - \frac{1}{4}, b + \frac{1}{4}, b + \frac{3}{4}$	$b^2 - \frac{1}{2}b + \frac{11}{144}$	$\frac{11}{144}$
$\frac{1}{2} < b < \frac{2}{3}$	$\frac{3}{2}b - \frac{1}{2}, \frac{1}{2}b + \frac{1}{2}$	$\frac{3}{4}b^2 - \frac{1}{2}b + \frac{5}{36}$	$\frac{1}{4}b^2 + \frac{5}{36}$
$b > \frac{2}{3}$	$\frac{1}{2}$	$\frac{5}{36}$	$b^2 + \frac{5}{36}$

C.3 The five-state case

Suppose that the set of states is $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}1\}$. As in the last subsection, we only report the ODS and losses for agents mentioned in the text. We report the analogous statistics for other amateurs in Anesi and Seidmann (2011).

The Table below describes the ODS and the loss incurred by the principal and the expert for generic bias:

b	Δ^*	λ^*	Λ^*
$\frac{1}{8} < b < \frac{1}{4}$	$\frac{4}{5}b - \frac{1}{10}, \frac{6}{5}b + \frac{1}{10}, \frac{4}{5}b + \frac{2}{5}, \frac{6}{5}b + \frac{3}{5}, \frac{4}{5}b + \frac{9}{10}$	$\frac{24}{25}b^2 - \frac{6}{25}b + \frac{3}{200}$	$\frac{1}{25}b^2 + \frac{3}{200}$
$\frac{1}{4} < b < \frac{3}{8}$	$\frac{6}{5}b - \frac{1}{5}, \frac{4}{5}b + \frac{1}{5}, \frac{6}{5}b + \frac{3}{10}, \frac{4}{5}b + \frac{7}{10}$	$\frac{24}{25}b^2 - \frac{8}{25}b + \frac{7}{200}$	$\frac{1}{25}b^2 + \frac{7}{200}$
$\frac{3}{8} < b < \frac{1}{2}$	$\frac{2}{5}b + \frac{1}{10}, \frac{8}{5}b - \frac{1}{10}, \frac{2}{5}b + \frac{3}{5}$	$\frac{16}{25}b^2 - \frac{7}{25}b + \frac{13}{200}$	$\frac{9}{25}b^2 + \frac{13}{200}$
$\frac{1}{2} < b < \frac{5}{8}$	$\frac{8}{5}b - \frac{1}{2}, \frac{2}{5}b + \frac{1}{2}$	$\frac{16}{25}b^2 - \frac{2}{5}b + \frac{1}{8}$	$\frac{9}{25}b^2 + \frac{1}{8}$
$b > \frac{5}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$b^2 + \frac{1}{8}$

If $E = \{0, 1\}$ then

b	Δ^{01}	λ^{01}	Λ^{01}
$b < \frac{7}{32}$	$\frac{1}{8}, \frac{2}{3}b + \frac{5}{12}, \frac{4}{3}b + \frac{7}{12}, \frac{2}{3}b + \frac{11}{12}$	$\frac{8}{15}b^2 - \frac{2}{15}b + \frac{7}{480}$	$\frac{7}{15}b^2 + \frac{7}{480}$
$\frac{7}{32} < b < \frac{1}{4}$	$\frac{4}{5}b - \frac{1}{20}, \frac{6}{5}b + \frac{3}{10}, \frac{4}{5}b + \frac{7}{10}, \frac{6}{5}b + \frac{4}{5}$	$\frac{24}{25}b^2 - \frac{8}{25}b + \frac{7}{200}$	$\frac{1}{25}b^2 + \frac{7}{200}$
$\frac{1}{4} < b < \frac{3}{8}$	$\frac{2}{5}b + \frac{1}{20}, \frac{8}{5}b + \frac{1}{5}, \frac{2}{5}b + \frac{4}{5}$	$\frac{16}{25}b^2 - \frac{6}{25}b + \frac{7}{200}$	$\frac{9}{25}b^2 + \frac{7}{200}$
$\frac{3}{8} < b < \frac{5}{8}$	$\frac{6}{5}b - \frac{1}{4}, \frac{4}{5}b + \frac{1}{2}$	$\frac{24}{25}b^2 - \frac{3}{5}b + \frac{1}{8}$	$\frac{1}{25}b^2 + \frac{1}{8}$
$b > \frac{5}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$b^2 + \frac{1}{8}$

If $E = \{12\}$ then

b	Δ^{12}	λ^{12}	Λ^{12}
$b < \frac{3}{16}$	$0, \frac{3}{8}, b + \frac{5}{8}, b + \frac{7}{8}$	$\frac{2}{5}b^2 - \frac{1}{10}b + \frac{1}{80}$	$\frac{3}{5}b^2 + \frac{1}{80}$
$\frac{3}{16} < b < \frac{1}{4}$	$\frac{4}{3}b - \frac{1}{4}, \frac{2}{3}b + \frac{1}{4}, b + \frac{5}{8}, b + \frac{7}{8}$	$\frac{14}{15}b^2 - \frac{3}{10}b + \frac{1}{32}$	$\frac{1}{15}b^2 + \frac{1}{32}$
$\frac{1}{4} < b < \frac{9}{32}$	$\frac{4}{3}b - \frac{1}{4}, \frac{2}{3}b + \frac{1}{4}, \frac{7}{8}$	$\frac{8}{15}b^2 - \frac{1}{5}b + \frac{1}{32}$	$\frac{7}{15}b^2 + \frac{1}{32}$
$\frac{9}{32} < b < \frac{11+\sqrt{265}}{48}$	$\frac{4}{5}b - \frac{1}{10}, \frac{6}{5}b + \frac{1}{10}, \frac{4}{5}b + \frac{13}{20}$	$\frac{24}{25}b^2 - \frac{11}{25}b + \frac{13}{200}$	$\frac{1}{25}b^2 + \frac{13}{200}$
$b > \frac{11+\sqrt{265}}{48}$	$\frac{1}{2}$	$\frac{1}{8}$	$b^2 + \frac{1}{8}$

If $E = \{2,3\}$ then

b	Δ^{23}	λ^{23}	Λ^{23}
$b < \frac{3}{16}$	$b - \frac{1}{8}, b + \frac{1}{8}, \frac{5}{8}, 1$	$\frac{2}{5}b^2 - \frac{1}{10}b + \frac{1}{80}$	$\frac{3}{5}b^2 + \frac{1}{80}$
$\frac{3}{16} < b < \frac{1}{3}$	$\frac{b}{2}, \frac{3}{2}b, \frac{1}{2}b + \frac{1}{2}, 1$	$\frac{3}{5}b^2 - \frac{1}{5}b + \frac{1}{40}$	$\frac{2}{5}b^2 + \frac{1}{40}$
$\frac{1}{3} < b < \frac{1}{2}$	$\frac{2}{5}b + \frac{1}{10}, \frac{8}{5}b - \frac{1}{10}, \frac{2}{5}b + \frac{3}{5}$	$\frac{16}{25}b^2 - \frac{7}{25}b + \frac{13}{200}$	$\frac{9}{25}b^2 + \frac{13}{200}$
$\frac{1}{2} < b < \frac{5}{8}$	$\frac{8}{5}b - \frac{1}{2}, \frac{2}{5}b + \frac{1}{2}$	$\frac{16}{25}b^2 - \frac{2}{5}b + \frac{1}{8}$	$\frac{9}{25}b^2 + \frac{1}{8}$
$b > \frac{5}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$b^2 + \frac{1}{8}$

If $E = \{0,1,2\}$ then

b	Δ^{012}	λ^{012}	Λ^{012}
$b < \frac{1}{4}$	$\frac{1}{4}, b + \frac{5}{8}, b + \frac{7}{8}$	$\frac{2}{5}b^2 - \frac{1}{10}b + \frac{1}{32}$	$\frac{3}{5}b^2 + \frac{1}{32}$
$\frac{1}{4} < b < \frac{5}{16}$	$\frac{1}{4}, \frac{7}{8}$	$\frac{1}{32}$	$b^2 + \frac{1}{32}$
$\frac{5}{16} < b < \frac{5}{8}$	$\frac{4}{5}b, \frac{6}{5}b + \frac{1}{2}$	$\frac{24}{25}b^2 - \frac{3}{5}b + \frac{1}{8}$	$\frac{1}{25}b^2 + \frac{1}{8}$
$b > \frac{5}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$b^2 + \frac{1}{8}$

If $E = \{1,2,3\}$ then

b	Δ^{123}	λ^{123}	Λ^{123}
$b < \frac{1}{4}$	$0, \frac{1}{2}, 1$	$\frac{1}{40}$	$b^2 + \frac{1}{40}$
$\frac{1}{4} < b < \frac{1}{2}$	$\frac{6}{5}b - \frac{3}{10}, \frac{4}{5}b + \frac{3}{10}, \frac{6}{5}b + \frac{7}{10}$	$\frac{24}{25}b^2 - \frac{12}{25}b + \frac{17}{200}$	$\frac{1}{25}b^2 + \frac{17}{200}$
$\frac{1}{2} < b < \frac{5}{8}$	$\frac{8}{5}b - \frac{1}{2}, \frac{2}{5}b + \frac{1}{2}$	$\frac{16}{25}b^2 - \frac{2}{5}b + \frac{1}{8}$	$\frac{9}{25}b^2 + \frac{1}{8}$
$b > \frac{5}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$b^2 + \frac{1}{8}$

If $E = \{0,1\},\{2,3\}$ then

b	$\Delta^{01,23}$	$\lambda^{01,23}$	$\Lambda^{01,23}$
$b < \frac{3}{16}$	$\frac{1}{8}, \frac{5}{8}, 1$	$\frac{1}{80}$	$b^2 + \frac{1}{80}$
$\frac{3}{16} < b < \frac{9}{32}$	$\frac{1}{8}, \frac{2}{3}b + \frac{1}{2}, \frac{4}{3}b + \frac{3}{4}$	$\frac{8}{15}b^2 - \frac{1}{5}b + \frac{1}{32}$	$\frac{7}{15}b^2 + \frac{1}{32}$
$\frac{9}{32} < b < \frac{11+\sqrt{265}}{48}$	$\frac{4}{5}b - \frac{1}{10}, \frac{6}{5}b + \frac{7}{20}, \frac{4}{5}b + \frac{9}{10}$	$\frac{24}{25}b^2 - \frac{11}{25}b + \frac{13}{200}$	$\frac{1}{25}b^2 + \frac{13}{200}$
$b > \frac{11+\sqrt{265}}{48}$	$\frac{1}{2}$	$\frac{1}{8}$	$b^2 + \frac{1}{8}$

If $E = \{0,2\},\{1,4\}$ then

b	$\Delta^{02,14,3}$	$\lambda^{02,14,3}$	$\Lambda^{02,14,3}$
$b < \frac{1}{16}$	$\frac{1}{4}, \frac{5}{8}, \frac{3}{4}$	$\frac{13}{160}$	$b^2 + \frac{13}{160}$
$\frac{1}{16} < b < \frac{1}{4}$	$\frac{1}{4}, \frac{2}{3}b + \frac{7}{12}, \frac{4}{3}b + \frac{2}{3}$	$\frac{8}{15}b^2 - \frac{1}{15}b + \frac{1}{12}$	$\frac{7}{15}b^2 + \frac{1}{12}$
$\frac{1}{4} < b < \frac{5}{8}$	$\frac{6}{5}b, \frac{4}{5}b + \frac{1}{2}$	$\frac{24}{25}b^2 - \frac{2}{5}b + \frac{1}{8}$	$\frac{1}{25}b^2 + \frac{1}{8}$
$b > \frac{5}{8}$	$\frac{1}{2}$	$\frac{1}{8}$	$b^2 + \frac{1}{8}$

References

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

Amador, M., I. Werning and G. Angeletos (2006), "Commitment vs. Flexibility" *Econometrica* 74, 365-396.

Ambrus, A. and G. Egorov (2009), "Delegation and Nonmonetary Incentives", mimeo.

Anesi, V. and D.J. Seidmann (2011), "Online Appendix to Accompany Optimal Delegation with a Finite Number of States."
http://www.nottingham.ac.uk/cedex/documents/papers/2011-04-appendix.pdf.

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

Athey, S. A. Atkeson and P. Kehoe (2005) "The Optimal Degree of Monetary Policy Discretion" *Econometrica* 73, 1431-1476.

Bendor, J. and A. Meirowitz (2004), "Spatial Models of Delegation" *American Political Science Review* 98, 293-310.

Callander, S. (2008), "A Theory of Policy Expertise" *Quarterly Journal of Political Science* 3, 123-140.

Crawford, V. and J. Sobel (1982), "Strategic Information Transmission" *Econometrica* 50, 1431-1451.

Dessein, W. (2002), "Authority and Communication in Organizations" *Review of Economic Studies* 69, 811-832.

Egorov, G. and K. Sonin (forthcoming), "Dictators and their Viziers: Endogenizing the Loyalty-Competence Trade-off" *Journal of the European Economic Association*

Fischer, P. and P. Stocken (2001), "Imperfect Information and Credible Communication" *Journal of Accounting Research* 39, 119-134.

Gailmard, S. and J. Patty (2007), "Slackers and Zealots: Civil Service, Policy Discretion, and Bureaucratic Expertise" *American Journal of Political Science* 51, 873-889.

Gailmard, S. (2009), "Discretion Rather Than Rules" *Political Analysis* 17, 25-44.

Goltsman, M., J. Hörner, G. Pavlov and F. Squintani (2009), "Mediation, Arbitration and Negotiation" *Journal of Economic Theory* 144, 1397-1420.

Green, J. (1982), "Statistical Decision Theory Requiring Incentives for Information Transfer" in J. McCall ed *The Economics of Information and Uncertainty* University of Chicago Press, Chicago.

Hannaford-Agor, P., V. Hans, N. Mott and G. Munsterman (2002) *Are Hung Juries a Problem?* National Institute of Justice, DC.

Harris, M. and A. Raviv (1996), "The Capital Budgeting Process, Incentives and Information" *Journal of Finance* 51, 1139-1174.

Hoffheimer, M. (2006), "The Future of Constitutionally Required Lesser Offenses" *University of Pittsburgh Law Review* 67, 585-640.

Holmstrom, B. (1984), "On the Theory of Delegation" in M. Boyer and R. Kihlstrom eds. *Bayesian Models in Economic Theory* Elsevier, Amsterdam.

Huber, J. and S. Gordon (2007), "Directing Retribution: On the Political Control of Lower Court Judges" *Journal of Law, Economics and Organization* 23, 386-420.

Huber, J. and N. McCarty (2004), "Bureaucratic Capacity, Delegation and Political Reform" *American Political Science Review* 98, 481-494.

Ivanov, M. (2010), "Informational Control and Organizational Design" *Journal of Economic Theory* 145, 721-751.

Kelman, S. and J. Meyers (2009), "Successfully Executing Ambitious Strategies in Government", mimeo.

Koessler, F. and D. Martimort (2009), "Optimal Delegation with Multidimensional Decisions", mimeo.

Kovac, E. and T. Mylovanov (2009), "Stochastic Mechanisms in Settings without Monetary Transfers: The Regular Case" *Journal of Economic Theory* 144, 1373-1395.

Krishna V. and J. Morgan (2008) "Contracting for Information under Imperfect Commitment" *Rand Journal of Economics* 39, 905-925.

Lewis, D. (2007) "Testing Pendleton's Premise: Do Political Appointees Make Worse Bureaucrats?" *Journal of Politics* 69, 1073-1088.

Luban, D. (1999), "Contrived Ignorance" *Georgetown Law Journal* 87, 957-980.

Martimort, D. and A. Semenov (2006), "Continuity in Mechanism Design without Transfers" *Economics Letters* 93, 182-189.

McCarty, N. (2004), "The Appointments Dilemma" *American Journal of Political Science* 48, 413-428.

Melumad, N. and T. Shibano (1991), "Communication in Settings with no Transfers" *Rand Journal of Economics* 22, 173-198.

Mylovanov, T. (2008), "Veto-Based Delegation" *Journal of Economic Theory* 138, 297-307.

Postlewaite, A. (1982), "Comment" in J. McCall ed *The Economics of Information and Uncertainty* University of Chicago Press, Chicago.

Szalay, D. (2005), "The Economics of Clear Advice and Extreme Opinions" *Review of Economic Studies* 72, 1173-1198.

Vickers, J. (1985), "Delegation and the Theory of the Firm" *Economic Journal* 95, 138-147.