

Statistical Reports for Remote Agents

Isaiah Andrews, *Harvard University and NBER**

Jesse M. Shapiro, *Brown University and NBER*

November 2018

Abstract

A statistician observes data informative about an unknown parameter and makes a report to an audience of agents who take a decision and suffer a loss. Agents may differ in their prior beliefs, information, or loss functions. We consider two notions of statistical risk for comparing reporting rules. The first, *decision risk*, evaluates the expected loss of an agent assuming the agent takes the decision prescribed by the report. The second, *remote risk*, evaluates the expected loss of an agent assuming the agent takes her optimal decision given the report. Rules that are appealing with respect to decision risk may be unappealing with respect to remote risk, and vice versa. For example, there is sometimes no rule that is admissible in both notions of risk, and hence no rule that minimizes weighted average risk with respect to full-support weights under both notions of risk. We obtain a more encouraging result for minimax optimality. Our results expose a possible tension in scientific research: rules that constitute good decisions (estimators) may not be good for conveying useful information to the audience, and vice versa.

*We acknowledge funding from the National Science Foundation under Grant No. 1654234, the Alfred P. Sloan Foundation, and the Silverman (1968) Family Career Development Chair at MIT. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding sources. We thank Glen Weyl for contributions to this project in its early stages. Conversations with Matthew Gentzkow and Kevin M. Murphy on related projects greatly influenced our thinking. For comments and helpful conversations, we also thank Alberto Abadie, Tim Armstrong, Gary Chamberlain, Max Kasy, Adam McCloskey, Emily Oster, Mikkel Plagborg-Møller, and Tomasz Strzalecki. We thank our dedicated research assistants for their contributions to this project. Our title references Hildreth (1963). E-mail: iandrews@fas.harvard.edu, jesse_shapiro_1@brown.edu.

1 Introduction

A traditional model of statistical research (e.g., Lehmann and Casella 1998) features settings in which a statistician observes data informative about an unknown parameter. The statistician then takes a decision or, equivalently, recommends one to an agent. The agent takes the recommended decision and suffers a loss that depends on the decision and on the parameter. From the standpoint of a given agent, a good statistical reporting rule (estimator) is one whose recommended decisions lead to low expected loss. Many standard criteria for choosing estimators (e.g., admissibility, efficiency) are (or can be) grounded in this model.

In many scientific settings, a statistician-scientist communicates a finding to a diverse audience that may include other researchers, practitioners, and policymakers. Different agents in the audience may have different prior opinions, information, or objectives (Hildreth 1963), which in turn may lead to different optimal decisions following a given statistical report. From the standpoint of a given agent, a good statistical reporting rule is one that provides the agent with information that is useful for making her decision, according to the agent's own beliefs and objectives.

Superficially, at least, these two types of settings are different. If good reporting rules under the first type of setting are not good reporting rules under the second, then the study of traditional decision-theoretic criteria may not be sufficient for understanding what statistics scientists should, or do, report.

In this paper, we ask whether the objective of recommending a good decision and the objective of providing useful information are compatible, in the sense that a single reporting rule can achieve both goals. To do this, we study a setting in which a statistician makes a report to an audience of agents. We consider two notions of the expected loss, or *risk*, for a given agent of a given reporting rule. The first, which we call *decision risk*, evaluates an agent's expected loss assuming the agent takes the decision prescribed by the report. This corresponds to the traditional Bayes risk (e.g., Lehmann and Casella 1998, Chapter 4). The second, which we call *remote risk*, evalu-

ates an agent's expected loss assuming the agent takes her optimal decision given the information conveyed by the report. This corresponds to the minimum of the agent's Bayes risk over decision rules that depend on the report, and has previously been considered by a number of authors including Hildreth (1963). We assume throughout that the sets of possible parameter values, data realizations, decisions, and reports are finite.

When there is a single agent, the distinction between the two notions of risk is not consequential, because the statistician can replace any reporting rule with a rule that prescribes the agent's preferred decision given the information in the report. When there are multiple agents, the distinction can be consequential, because agents may have different optimal decisions given the information in the report.

Indeed, we find that the conflict between the two notions of risk can be severe. A rule is admissible with respect to a given notion of risk if it is not dominated by any other rule, in the sense that another rule yields weakly lower risk for all agents in the audience and strictly lower risk for at least one. We show that if (i) the set of feasible decisions is smaller than the set of distinct optimal action profiles for agents and (ii) some decision is dominated in loss, then any rule that is admissible in decision risk is inadmissible (i.e., dominated) in remote risk, and vice versa. We develop examples, based on contexts of practical interest, in which conditions (i) and (ii) are satisfied.

A similar conflict arises when considering rules that minimize weighted average risk with respect to full-support weights over the audience. We show that such rules are admissible. Hence, if there is no rule that is admissible in both decision and remote risk, it follows that any rule that minimizes weighted average decision risk does not minimize weighted average remote risk with respect to *any* full-support weights, and vice-versa. That is, under full support, there exists no Bayes estimator that minimizes weighted average remote risk, no matter which weights we choose for the audience.

We obtain a more encouraging result for rules that minimize the maximum risk over the agents in the audience (i.e., minimax rules). If the set of beliefs held by the

audience is convex, then there exist rules that are minimax in both decision risk and remote risk. Our other results imply that such rules will sometimes be inadmissible, so this result reconciles the two notions of risk only if minimaxity is truly the criterion of interest.

Our results suggest that, in many cases, a scientist wishing to convey useful information to a diverse audience may choose an estimator that does not constitute a good decision rule. In our conclusion, we discuss possible implications for scientific practice.

We are not aware of past work that compares the recommendations of decision and remote risk in a setting with heterogeneous agents. Raiffa and Schlaifer (1961), Hildreth (1963), Sims (1982, 2007), and Geweke (1997, 1999), among others, consider the problem of communicating statistical findings to diverse, Bayesian agents.¹ Our analysis is particularly related to that of Hildreth (1963) who studied, among other topics, the properties of what we term remote risk in the single-agent setting.

Our setting is also related to the literature on comparisons of experiments following Blackwell (1951, 1953).² What we term remote risk has previously appeared in this literature (see for instance example 1.4.5 in Torgersen 1991), but the primary focus has been on properties (e.g., Blackwell's order) that hold for all possible beliefs and loss functions. By contrast, we follow the literature on statistical decision theory (e.g., Lehmann and Casella 1998) and consider a given loss function and class of priors. Our analysis focuses on the comparison between remote and decision risk.

Our setting is broadly related to large literatures on strategic communication (Crawford and Sobel 1982) and information design (Bergemann and Morris forthcoming). As in Farrell and Gibbons (1989), the receivers (agents) in our setting are heterogeneous. As in Kamenica and Gentzkow (2011), the sender (statistician) in our

¹See also Poirier (1988) and Andrews et al. (2017). A related literature (e.g., Pratt 1965, Kwan 1999, Abadie 2018, Abadie and Kasy forthcoming, Frankel and Kasy 2018) assesses the Bayesian interpretation of frequentist inference.

²Le Cam (1996) provides a brief review, while an extensive treatment can be found in Torgersen (1991).

setting commits in advance to a reporting strategy. Unlike much of the literature on strategic communication, our setting does not involve a conflict of interest between the sender and the receivers.

The remainder of the paper is organized as follows. Section 2 introduces our setting, notation, and key definitions, and provides some preliminary results. Section 3 considers choosing rules based on admissibility. Section 4 considers choosing rules based on weighted average risk. Section 5 considers choosing rules based on maximum risk. Section 6 concludes and discusses possible implications for scientific practice. Appendix A provides proofs, while Appendix B discusses auxiliary results.

2 Setting

2.1 Primitives

A statistician observes data $X \in \mathcal{X}$ for \mathcal{X} a finite sample space, $|\mathcal{X}| < \infty$. The distribution of X is governed by a parameter $\theta \in \Theta$, with $X|\theta \sim F_\theta$, for Θ a finite parameter space. We assume that F_θ has support equal to \mathcal{X} for all $\theta \in \Theta$, and refer to $\{F_\theta : \theta \in \Theta\}$ as the *model*. The statistician also observes a public random variable $V \sim U[0, 1]$ that is independent of θ and X , and so serves as a public randomization device.

After observing (X, V) , the statistician makes a report $s \in \mathcal{S}$ for \mathcal{S} a finite signal space. This report, along with the the public random variable V , is observed by every agent a in an audience \mathcal{A} . Agents differ only in their priors on θ , so without loss of generality we identify an agent with her prior.³ Hence, $\mathcal{A} \subseteq \Delta(\Theta)$ for $\Delta(\Theta)$ the set of all distributions on Θ . We assume that \mathcal{A} is closed.

After observing the statistician's report s and the public random variable V , each agent takes a decision $d \in \mathcal{D}$ for \mathcal{D} a finite decision space, and then realizes a loss

³In the few cases where we need to distinguish between an agent and her prior, we denote the agent by a and the prior by $a(\cdot)$.

$L(d, \theta)$, where $0 \leq L(d, \theta) < \infty$ for all $d \in \mathcal{D}, \theta \in \Theta$. We assume throughout that $\mathcal{D} \subseteq \mathcal{S}$, so the signal space is sufficiently rich for the statistician to recommend any feasible decision to the audience.

Remark 1. Interpretation of the audience. We have described the setting as one with an audience of agents who can differ in their prior beliefs (Morris 1995). We can equivalently view the setting as one with a single agent who, after receiving the statistician’s report but before making her decision, receives a piece of information a that is informative about θ . Appendix B.1 shows that settings where an audience of agents share a common prior but have heterogeneous loss functions, or where both priors and loss functions are heterogeneous, can also be cast into our setting through appropriate relabeling. Brown (1975) considers a related setting with a collection of possible loss functions and proposes notions of admissibility for that setting.

Remark 2. Disagreement about the model. Our assumption that all disagreement between agents can be expressed as differences in the prior allows for the possibility that agents disagree about the model. Specifically, suppose that we have K separate audiences $\mathcal{A}_1, \dots, \mathcal{A}_K$ where agents in audience k have priors on the finite parameter space Θ_k , and believe in model k which implies that conditional on true parameter value $\theta_k \in \Theta_k$, $X \sim F_{k, \theta_k}$. If we then define $\Theta = \cup_k \cup_{\theta_k \in \Theta_k} (k, \theta_k)$ and $F_\theta = F_{k, \theta_k}$, disagreement about the model can be expressed as heterogeneity in prior beliefs defined on the enriched parameter space.

Remark 3. Interpretation of the communication constraint. Finiteness of the signal space \mathcal{S} is an important assumption. It encodes limits on communication between the statistician and the audience similar to the information-theoretic metaphor of a code with a finite alphabet (Cover and Thomas 2006). Such limits on communication could arise either from external constraints (e.g., data privacy) or from limits on the audience’s ability to process information. Common scientific practices such as publishing summaries of data (and not merely the data themselves) suggest that such limits are relevant in scientific research. Our assumption that $\mathcal{D} \subseteq \mathcal{S}$ then implies

that the decision space is finite as well.

Remark 4. Assumption of finite sample and parameter spaces. In addition to the signal space \mathcal{S} and the decision space \mathcal{D} , we also assume that the sample space \mathcal{X} and the parameter space Θ are finite, though the audience \mathcal{A} need not be. Finiteness of \mathcal{X} and Θ is technically convenient, allowing us to avoid measure-theoretic complications, and similar assumptions appear in related work (Blackwell 1951, 1953; Grünwald and Dawid 2004; Müller and Norets 2016). All information in statistics is finite in practice — all measurements of the world are taken to finite precision — but a major drawback of our assumptions is that they rule out many standard statistical settings (e.g., estimating the mean of a normal random variable).

Example 1. To make our setting more concrete, consider an example with $|\Theta| > 1$, $\Theta \subset \mathcal{D} = \mathcal{S} \subseteq \mathcal{X} \subset \mathbb{R}$, $\mathcal{A} = \Delta(\Theta)$, and $L(d, \theta) = l(|d - \theta|)$ for $l(\cdot)$ strictly convex and minimized at 0. This can be viewed as a discrete version of the framework employed in many estimation problems, where convex loss functions (for example quadratic loss) are often considered. \blacktriangle

2.2 Risk

The statistician publicly commits to a rule $c : \mathcal{X} \times [0, 1] \rightarrow \Delta(\mathcal{S})$ that maps from the data X and the public random variable V to a distribution supported on the signal space \mathcal{S} . We denote by C the random variable equal to the report generated by c , so that $C|X, V \sim c(X, V)$.⁴

We denote by $\mathcal{C}_{\mathcal{S}'}$ the set of all rules $c : \mathcal{X} \times [0, 1] \rightarrow \Delta(\mathcal{S}')$ that map to distributions supported on the nonempty set $\mathcal{S}' \subseteq \mathcal{S}$, so that $\mathcal{C}_{\mathcal{D}}$ is the set of rules with reports supported on \mathcal{D} , and $\mathcal{C}_{\mathcal{S}}$ (abbreviated by \mathcal{C}) is the set of all rules. We denote by \mathcal{B} the set of “binding” rules $b : \mathcal{S} \times [0, 1] \rightarrow \Delta(\mathcal{D})$ that map from the signal and

⁴The public random variable V allows the statistician to randomize among a set of deterministic rules $c : \mathcal{X} \rightarrow \mathcal{S}$ while also informing the audience of which deterministic rule was selected. Privately randomized rules $c : \mathcal{X} \rightarrow \Delta(\mathcal{S})$ can be viewed as randomizing among a class of deterministic rules without telling the audience which deterministic rule was selected.

the public random variable to a distribution over decisions, and write $b \circ c \in \mathcal{C}_{\mathcal{D}}$ for the composition of the rules $c \in \mathcal{C}$ and $b \in \mathcal{B}$.

Definition 1. The **decision risk** $R_a(c)$ of a rule $c \in \mathcal{C}_{\mathcal{D}}$ for an agent a is

$$R_a(c) = \mathbb{E}_a [L(C, \theta)]$$

where, here and throughout, \mathbb{E}_a denotes the expectation with respect to agent a 's beliefs.

The **remote risk** $R_a^*(c)$ of a rule $c \in \mathcal{C}$ for an agent a is

$$R_a^*(c) = \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] \right].$$

The decision risk of a rule $c \in \mathcal{C}_{\mathcal{D}}$ for agent $a \in \mathcal{A}$ is simply the rule's average risk with respect to the prior $a(\cdot)$. This is also known as the Bayes risk with respect to $a(\cdot)$; see, e.g., Chapter 4 of Lehmann and Casella (1998). A rule that minimizes decision risk for a is thus a Bayes decision rule with respect to $a(\cdot)$.

In contrast to decision risk, which evaluates the expected loss assuming every member of the audience takes the action prescribed by the report, remote risk allows agents to take different actions in response to C . The remote risk is simply the minimum Bayes risk over decision rules that depend on (C, V) ,

$$R_a^*(c) = \min_{b \in \mathcal{B}} R_a(b \circ c).$$

For any rule $c \in \mathcal{C}_{\mathcal{D}}$ and any agent $a \in \mathcal{A}$, decision risk exceeds remote risk, i.e. $R_a^*(c) \leq R_a(c)$, and this inequality may be strict. For any rule $c \in \mathcal{C}$ and any agent $a \in \mathcal{A}$, there always exists a rule $\tilde{c}_a \in \mathcal{C}_{\mathcal{D}}$ such that $R_a^*(c) = R_a(\tilde{c}_a)$, since we can take \tilde{c}_a to be a 's optimal action after observing X (see, for example, Bergemann and Morris forthcoming). By contrast, there may exist no rule $\tilde{c} \in \mathcal{C}_{\mathcal{D}}$ such that $R_a^*(c) = R_a(\tilde{c})$ for all $a \in \mathcal{A}$ (consider for instance the subset of the agents in Example 1 who have

dogmatic priors). In this sense the distinction between decision and remote risk is inconsequential when there is a single agent a but may be consequential when there are multiple agents.

Example 1 (Continued) If in this example we consider quadratic loss $l(x) = x^2$, the decision risk $R_a(c)$ is simply the mean squared error for C averaged over a 's prior $\sum_{\theta \in \Theta} a(\theta) \mathbb{E}_\theta [(C - \theta)^2]$. By contrast, the remote risk $R_a^*(c)$ is the average squared error of a 's optimal prediction for θ after observing C , $\mathbb{E}_a [\min_{d \in \mathcal{D}} \mathbb{E}_a [(\theta - d)^2 | C, V]]$. If, for some $a \in \mathcal{A}$, the set \mathcal{D} always contains a 's posterior mean for θ , i.e., $\mathbb{E}_a [\theta | C, V] \in \mathcal{D}$ for all C, V , then $R_a^*(c)$ simplifies to a 's average posterior variance $\mathbb{E}_a [\text{Var}_a(\theta | C, V)]$. \blacktriangle

2.3 Preliminaries

For some of our analysis, it will be convenient to restrict attention to rules such that C is non-random conditional on (X, V) . There is no increase in risk from imposing this restriction.

Lemma 1. *For any $c \in \mathcal{C}_{\mathcal{D}}$, there exists a non-randomized rule $\tilde{c} : \mathcal{X} \times [0, 1] \rightarrow \mathcal{D}$ such that $R_a(\tilde{c}) = R_a(c)$ for all $a \in \mathcal{A}$. Similarly, for any nonempty $\mathcal{S}' \subseteq \mathcal{S}$ and $c \in \mathcal{C}_{\mathcal{S}'}$, there exists a non-randomized rule $\tilde{c} : \mathcal{X} \times [0, 1] \rightarrow \mathcal{S}'$ such that $R_a^*(\tilde{c}) \leq R_a^*(c)$ for all $a \in \mathcal{A}$. Denote the class of such rules by $\mathcal{C}_{\mathcal{S}'}^N$.*

The proof of Lemma 1 shows that for any c we can construct a rule \tilde{c} such that (i) \tilde{C} is nonrandom conditional on (X, V) and (ii) \tilde{C} has the same conditional distribution given X as C . This immediately implies that the rule \tilde{c} yields the same decision risk as c , and in the proof we show that \tilde{c} may yield strictly lower remote risk than c . Intuitively, this is true because randomization conditional on (X, V) adds unnecessary noise to the signal. Viewing V as a public randomization device, we thus see that public and private randomization generate the same decision risk but may imply different remote risk.

Definition 2. For some $\varepsilon \geq 0$, the rule $c \in \mathcal{C}_{\mathcal{D}}$ is ε -**dominated in decision risk** by rule $c^* \in \mathcal{C}_{\mathcal{D}}$ if $R_a(c) - \varepsilon \geq R_a(c^*)$ for all $a \in \mathcal{A}$. The rule $c \in \mathcal{C}_{\mathcal{D}}$ is ε -**admissible in decision risk** if it is not ε -dominated in decision risk by any rule $c^* \in \mathcal{C}_{\mathcal{D}}$. The rule $c \in \mathcal{C}_{\mathcal{D}}$ is **extended admissible in decision risk** if it is ε -admissible in decision risk for all $\varepsilon > 0$.

For some $\varepsilon \geq 0$, the rule $c \in \mathcal{C}$ is ε -**dominated in remote risk** by rule $c^* \in \mathcal{C}$ if $R_a^*(c) - \varepsilon \geq R_a^*(c^*)$ for all $a \in \mathcal{A}$. For nonempty $\mathcal{S}' \subseteq \mathcal{S}$, the rule $c \in \mathcal{C}_{\mathcal{S}'}$ is ε -**admissible in remote risk** with respect to \mathcal{S}' if it is not ε -dominated in remote risk by any rule $c^* \in \mathcal{C}_{\mathcal{S}'}$. The rule $c \in \mathcal{C}_{\mathcal{S}'}$ is **extended admissible in remote risk** with respect to \mathcal{S}' if it is ε -admissible in remote risk with respect to \mathcal{S}' for all $\varepsilon > 0$.

Proposition 1. There exists a rule $c \in \mathcal{C}_{\mathcal{D}}$ that is both extended admissible in decision risk and extended admissible in remote risk with respect to \mathcal{S} .

The proof of Proposition 1 uses the fact that a rule $c \in \mathcal{C}_{\mathcal{D}}$ that minimizes decision risk for any agent a is necessarily extended admissible in both decision risk and remote risk. As this fact suggests, extended admissibility is a weak requirement. For example, if as in Example 1 the audience \mathcal{A} contains an agent with a dogmatic prior, then all rules $c \in \mathcal{C}$ are extended admissible in remote risk with respect to \mathcal{S} , since no signal leads this agent to update. We therefore view Proposition 1 as a point of departure for our analysis, rather than as a substantive reconciliation of the two notions of risk.

In the following sections we ask if it is possible to reconcile the recommendations of remote and decision risk under three notions of optimality that are stronger than extended admissibility: admissibility, optimality in weighted average risk, and minimaxity.

3 Admissibility

Definition 3. The rule $c \in \mathcal{C}_{\mathcal{D}}$ is **dominated in decision risk** by rule $c^* \in \mathcal{C}_{\mathcal{D}}$ if $R_a(c) \geq R_a(c^*)$ for all $a \in \mathcal{A}$, with strict inequality for at least one $a \in \mathcal{A}$. The rule $c \in \mathcal{C}_{\mathcal{D}}$ is **admissible in decision risk** if it is not dominated in decision risk by any rule $c^* \in \mathcal{C}_{\mathcal{D}}$.

The rule $c \in \mathcal{C}$ is **dominated in remote risk** by rule $c^* \in \mathcal{C}$ if $R_a^*(c) \geq R_a^*(c^*)$ for all $a \in \mathcal{A}$, with strict inequality for at least one $a \in \mathcal{A}$. For nonempty $\mathcal{S}' \subseteq \mathcal{S}$, the rule $c \in \mathcal{C}_{\mathcal{S}'}$ is **admissible in remote risk** with respect to \mathcal{S}' if it is not dominated in remote risk by any rule $c^* \in \mathcal{C}_{\mathcal{S}'}$.

Dominance in decision risk is closely related to conventional decision-theoretic notions of dominance. If $\mathcal{A} = \Delta(\Theta)$, as in Example 1, then a rule $c \in \mathcal{C}_{\mathcal{D}}$ is admissible in decision risk if and only if it is an admissible decision rule in the usual sense (e.g., Lehmann and Casella 1998, p. 48).

Dominance in remote risk is related to Blackwell’s order. If we consider rules c and \tilde{c} such that the induced distribution of (C, V) is more informative about θ than that of (\tilde{C}, V) in the sense of Blackwell (1953), then c has weakly smaller remote risk than \tilde{c} for all $a \in \mathcal{A}$. In particular, if C is a sufficient statistic for θ , then c is admissible in remote risk with respect to \mathcal{S} , and achieves weakly lower remote risk than any alternative rule for all agents.

We next ask whether we can always find a rule $c \in \mathcal{C}_{\mathcal{D}}$ that is admissible in decision risk and is admissible in remote risk with respect to \mathcal{D} . As admissibility is a weak requirement for estimators,⁵ failure to find such a rule indicates that any estimator $c \in \mathcal{C}_{\mathcal{D}}$ that is reasonable from the perspective of decision risk “leaves value on the table” from the perspective of remote risk, in the sense that there is another rule that would improve some agent’s expected payoff without worsening any other agent’s

⁵Lehmann and Casella (1998, p. 337) write that “...admissibility is an extremely weak property. While it is somewhat embarrassing for an estimator to be inadmissible, the fact that it is admissible in no way guarantees that it is a good or even halfway reasonable estimator.”

expected payoff.

Our question is interesting only if the set of rules that are admissible in decision risk and the set of rules that are admissible in remote risk are both nonempty. We verify that this is the case.

Lemma 2. *There is at least one rule $c \in \mathcal{C}_{\mathcal{D}}$ that is admissible in decision risk. For all nonempty $\mathcal{S}' \subseteq \mathcal{S}$, there is at least one rule $c \in \mathcal{C}_{\mathcal{S}'}$ that is admissible in remote risk with respect to \mathcal{S}' .*

Although the set of rules that are admissible in decision risk and the set of rules that are admissible in remote risk are both nonempty, for a given signal space these sets can fail to intersect. We give sufficient conditions for this to occur.

Definition 4. *Let \mathcal{P} be the set of partitions of \mathcal{X} , with generic element $P \in \mathcal{P}$. Let \mathcal{P}^* denote the subset of \mathcal{P} such that for every cell $\mathcal{X}_p \in P \in \mathcal{P}^*$, each agent has at least one decision $d \in \mathcal{D}$ that is optimal for every $X \in \mathcal{X}_p$. That is,*

$$\mathcal{P}^* = \left\{ P \in \mathcal{P} : \left\{ \bigcap_{X \in \mathcal{X}_p} \arg \min_{d \in \mathcal{D}} E_a [L(d, \theta) | X] \right\} \neq \emptyset \text{ for all } \mathcal{X}_p \in P, a \in \mathcal{A} \right\}.$$

The **effective size of the sample space** \mathcal{X} for the audience \mathcal{A} , denoted $N(\mathcal{X}, \mathcal{A})$, is the minimal size of a partition in \mathcal{P}^*

$$N(\mathcal{X}, \mathcal{A}) = \min \{|P| : P \in \mathcal{P}^*\}.$$

The effective size of the sample space is the size of the smallest partition we can make of the sample space \mathcal{X} such that knowing only which cell contains X , rather than X itself, still allows all agents to take optimal decisions. For example, such a partition would group redundant data realizations X and X' that imply the same likelihood under the model. The effective size of the sample space depends on the audience \mathcal{A} , and is increasing in \mathcal{A} . To see why, consider that if there is a single agent then $N(\mathcal{X}, \mathcal{A}) \leq |\mathcal{D}|$, whereas this need not hold with multiple agents.

The effective size of the sample space $N(\mathcal{X}, \mathcal{A})$ characterizes the point at which the communication constraint binds, in the sense that any reduction in the size of the signal space below $N(\mathcal{X}, \mathcal{A})$ leads to a strict increase in remote risk for some agent.

Lemma 3. For any nonempty $\mathcal{S}' \subseteq \mathcal{S}$ and any nonempty $\mathcal{S}'' \subset \mathcal{S}'$, there exists a rule $c^* \in \mathcal{C}_{\mathcal{S}''}$ that is admissible in remote risk with respect to \mathcal{S}' if and only if $|\mathcal{S}''| \geq N(\mathcal{X}, \mathcal{A})$.

The proof of Lemma 3 is intuitive. If $|\mathcal{S}''| \geq N(\mathcal{X}, \mathcal{A})$, then there exists a partition P of \mathcal{X} with $|P| \leq |\mathcal{S}''|$ such that knowing which cell in P contains the data allows all agents to obtain the same risk as if they had observed the full data. Hence, there is no loss in remote risk from limiting the signal space to \mathcal{S}'' . By contrast, when $|\mathcal{S}''| < N(\mathcal{X}, \mathcal{A})$, for any $c \in \mathcal{C}_{\mathcal{S}''}$ there exists $\mathcal{X}^* \subseteq \mathcal{X}$ such that (i) different elements of \mathcal{X}^* would lead some agent a to take different actions and (ii) c sometimes assigns the same signal to all $X \in \mathcal{X}^*$. Enlarging the signal space from \mathcal{S}'' to \mathcal{S}' allows us to assign a unique signal to at least one element of \mathcal{X}^* , which strictly reduces remote risk for a without increasing it for any other agent. Hence, the rule c is dominated in remote risk with respect to \mathcal{S}' .

Example 1 (Continued) Suppose that the model has monotone likelihood ratios in the sense that for all $\theta' > \theta$, $X' > X$,

$$\frac{dF_{\theta'}(X')}{dF_{\theta}(X')} > \frac{dF_{\theta'}(X)}{dF_{\theta}(X)}.$$

Claim 1. $N(\mathcal{X}, \mathcal{A}) = |\mathcal{X}|$ in Example 1.

Intuitively, the claim holds because the monotone likelihood ratio property ensures that agents' posteriors are increasing in X , and we have assumed a sufficient degree of prior heterogeneity ($\mathcal{A} = \Delta(\Theta)$) so that for any distinct (X, X') there is some agent whose optimal action differs between X and X' . \blacktriangle

Example 2. Let $|\Theta| > 1$, $\mathcal{D} = \mathcal{S} = \mathcal{X} = \Theta \cup \{\iota\}$, $\mathcal{A} = \Delta(\Theta)$, and $L(d, \theta) = \mathbf{1}_{d \neq \theta}$. Suppose that there are $\delta, \gamma \in (0, \frac{1}{4})$ such that for each $\theta \in \Theta$, (i) $\Pr_{\theta}\{X = \theta\} = 1 - \delta - \gamma$, (ii) $\Pr_{\theta}\{X = \theta'\} = \delta / (|\Theta| - 1)$ for each $\theta' \in \Theta \setminus \{\theta\}$, and (iii) $\Pr_{\theta}\{X = \iota\} = \gamma$, where $\Pr_{\theta}\{\cdot\}$ denotes the probability under F_{θ} .

In this example we have either that $X \in \Theta$, which suggests that $\theta = X$, or that $X = \iota$, which can be interpreted as a realization of the data that is uninformative about the parameter. The 0 – 1 loss function reflects the goal of choosing the single best estimate or hypothesis from among a finite set.

Claim 2. $N(\mathcal{X}, \mathcal{A}) = |\mathcal{X}|$ in Example 2.

Intuitively, the claim holds because the signal X is either informative about the parameter value ($X \in \Theta$), in which case some agents want to set $d = X$, or the signal is uninformative ($X = \iota$), in which case all agents want to set d equal to their prior mode, which differs across agents. \blacktriangle

Definition 5. A decision $d \in \mathcal{D}$ is **dominated in loss** if there exists $d' \in \mathcal{D}$ such that $L(d, \theta) \geq L(d', \theta)$ for all $\theta \in \Theta$, with strict inequality for some θ' such that $\Pr_{a'}\{\theta = \theta'\} > 0$ for at least one agent $a' \in \mathcal{A}$.

A decision $d \in \mathcal{D}$ is dominated in loss if another decision yields a weakly smaller loss under all parameter values, with a strictly smaller loss under a parameter value that some agent thinks is possible.

Lemma 4. If $d \in \mathcal{D}$ is dominated in loss, any rule $c \in \mathcal{C}_{\mathcal{D}}$ such that $\Pr\{C = d|X\} > 0$ for some $X \in \mathcal{X}$ is inadmissible in decision risk.

The probability $\Pr\{C = d|X\}$ is taken without reference to an agent a because all agents agree on the distribution of C conditional on X .

Proposition 2. If $N(\mathcal{X}, \mathcal{A}) \geq |\mathcal{D}|$ and there exists $d \in \mathcal{D}$ that is dominated in loss, then any rule $c \in \mathcal{C}_{\mathcal{D}}$ that is admissible in decision risk is not admissible in remote risk with respect to \mathcal{D} , and vice versa.

By Lemma 3, $N(\mathcal{X}, \mathcal{A}) \geq |\mathcal{D}|$ implies that rules admissible in remote risk with respect to \mathcal{D} always use the full decision space to communicate. On the other hand, Lemma 4 implies that rules admissible in decision risk never use the dominated decision d . Hence, there is necessarily a conflict between admissibility in remote and decision risk. Note that any rule c that is inadmissible in remote risk with respect to some nonempty $\mathcal{S}'' \subseteq \mathcal{S}'$ is inadmissible in remote risk with respect to \mathcal{S}' as well. Consequently, a conflict between admissibility in remote and decision risk arises even more readily when we contemplate use of the full signal space, rather than restricting reports to lie in \mathcal{D} .

Example 1 (Continued) In this example, suppose further that

$$\max \{\mathcal{D}\} > \max \{\Theta\},$$

reflecting a case in which we have *a priori* restrictions on the parameter space, for instance because a theoretical model implies that some parameters must be negative.

Claim 3. Any $d \in \mathcal{D}$ with $d > \max \{\Theta\}$ is dominated in loss in Example 1.

Intuitively, a decision d with $d > \max \{\Theta\}$ is dominated in loss because with a convex loss function agents strictly prefer a decision on the boundary of the parameter space to one outside of the parameter space, and a decision on the boundary of the parameter space is feasible since $\Theta \subset \mathcal{D}$. At the same time, censoring the report to the boundary of the parameter space sacrifices useful information. Hence, Proposition 2 applies to this example, and no rule is admissible in both decision risk and remote risk. \blacktriangle

Example 2 (Continued)

Claim 4. The decision $d = \iota$ is dominated in loss in Example 2.

Intuitively, $d = \iota$ is dominated in loss because all agents would weakly prefer any other $d' \in \mathcal{D}$, with some agents strictly preferring each such d' . At the same

time, from the perspective of remote risk it is useful to report when the data are uninformative about θ . Hence, Proposition 2 again applies, and no rule is admissible in both decision risk and remote risk. \blacktriangle

Remark 5. Admissibility conflict under a binding communication constraint. Consider a setting where $N(\mathcal{X}, \mathcal{A}) > |\mathcal{D}|$. If we augment the decision space with an element d_L that is dominated in loss, we immediately obtain an admissibility conflict in the sense of Proposition 2. In this sense it is easy to arrive at an admissibility conflict once the communication constraint is binding over the decision space.

Remark 6. Sufficient conditions for lack of conflict. One can likewise give a variety of sufficient conditions for remote and decision admissibility not to be in conflict. Perhaps the simplest sufficient condition is that there is a single agent, i.e., $\mathcal{A} = \{a\}$, since in this case setting c equal to an optimal decision rule for agent a , so that $R_a(c) = \min_{\tilde{c} \in \mathcal{C}_{\mathcal{D}}} R_a(\tilde{c})$, ensures optimality in both remote and decision risk. Another sufficient condition is that there exists a sufficient statistic for θ that is also an admissible decision rule.

Remark 7. Reporting bounds. One approach to addressing prior heterogeneity is to report bounds containing the optimal decision over a wide class of different priors (see, e.g., Chamberlain and Leamer 1976, Leamer 1982, Giacomini and Kitagawa 2015). A modification of Example 2 illustrates a context in which such procedures can be unappealing from the standpoint of remote risk. Suppose that $\mathcal{S} = 2^{\mathcal{D}}$ for $2^{\mathcal{D}}$ the power set of \mathcal{D} . Let c^B be the rule whose report C^B satisfies, for each X ,

$$C^B = \left\{ \bigcup_{a \in \mathcal{A}} \arg \min_{d \in \mathcal{D}} E_a [L(d, \theta) | X] \right\}.$$

Then the rule c^B is inadmissible in remote risk with respect to \mathcal{S} . It is dominated in

remote risk by the rule c^* whose report C^* satisfies

$$C^* = \begin{cases} \{X\}, & X \in \Theta \\ \mathcal{D}, & X = \iota \end{cases}.$$

4 Weighted Average Risk

The set of admissible rules is large in many cases, and admissibility provides no guidance for choosing a rule from this set. To obtain a sharper ranking over rules, a natural possibility is to consider a weighted average of risk over the audience. Let Ω be the set of weights (i.e., probability measures) ω supported on \mathcal{A} . Let $\Omega_+ \subset \Omega$ be the set of weights ω with support equal to \mathcal{A} .

Definition 6. *The **weighted average decision risk** of a rule $c \in \mathcal{C}_{\mathcal{D}}$ with respect to weights $\omega \in \Omega$ is $\int_{\mathcal{A}} R_a(c) d\omega(a)$.*

*The **weighted average remote risk** of a rule $c \in \mathcal{C}_{\mathcal{D}}$ with respect to weights $\omega \in \Omega$ is $\int_{\mathcal{A}} R_a^*(c) d\omega(a)$.*

Note that

$$\int_{\mathcal{A}} R_a(c) d\omega(a) = \sum_{\theta \in \Theta} \mathbb{E}_{\theta} [L(C, \theta)] \int_{\mathcal{A}} a(\theta) d\omega(a), \quad (1)$$

so the weighted average decision risk is equal to the classical average risk (see, e.g., Chapter 4 of Lehmann and Casella 1998) based on the weighted average prior $a_{\omega}(\theta) = \int a(\theta) d\omega(a)$. Hence, for decision risk we can collapse the heterogeneity in the audience to obtain an equivalent single-agent problem.

By contrast,

$$\int_{\mathcal{A}} R_a^*(c) d\omega(a) = \int_{\mathcal{A}} \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] \right] d\omega(a), \quad (2)$$

which is in general different than the remote risk for the weighted average prior

$a_\omega(\theta)$.⁶ Hence, we cannot in general express the weighted average remote risk as the remote risk for a single agent with the weighted average prior.

We can rank procedures in terms of their weighted average risk.

Definition 7. A rule $c \in \mathcal{C}_\mathcal{D}$ is **optimal in weighted average decision risk** with respect to weights $\omega \in \Omega$ if

$$\int_{\mathcal{A}} R_a(c) d\omega(a) = \inf_{\tilde{c} \in \mathcal{C}_\mathcal{D}} \int_{\mathcal{A}} R_a(\tilde{c}) d\omega(a). \quad (3)$$

For nonempty $\mathcal{S}' \subseteq \mathcal{S}$, a rule $c \in \mathcal{C}_{\mathcal{S}'}$ is **optimal in weighted average remote risk** with respect to weights $\omega \in \Omega$ and signal space \mathcal{S}' if

$$\int_{\mathcal{A}} R_a^*(c) d\omega(a) = \inf_{\tilde{c} \in \mathcal{C}_{\mathcal{S}'}} \int_{\mathcal{A}} R_a^*(\tilde{c}) d\omega(a). \quad (4)$$

There always exists a weighted average risk optimal rule.

Lemma 5. For any weights $\omega \in \Omega$, there exists a rule $c \in \mathcal{C}_\mathcal{D}$ that is optimal in weighted average decision risk with respect to weights ω . For any weights $\omega \in \Omega$ and nonempty $\mathcal{S}' \subseteq \mathcal{S}$, there exists a rule c that is optimal in weighted average remote risk with respect to weights ω and signal space \mathcal{S}' .

Classical complete class theorems for decision risk say that by varying the weights ω we can trace out the full set of admissible rules. This result continues to hold for remote risk.

Definition 8. A set of rules $\mathcal{C}^* \subseteq \mathcal{C}_\mathcal{D}$ is **complete in decision risk** if any rule $c \in \mathcal{C}_\mathcal{D} \setminus \mathcal{C}^*$ is dominated in decision risk by some rule $c^* \in \mathcal{C}^*$. For nonempty $\mathcal{S}' \subseteq \mathcal{S}$, a set of rules $\mathcal{C}^* \subseteq \mathcal{C}_{\mathcal{S}'}$ is **complete in remote risk** with respect to \mathcal{S}' if any rule $c \in \mathcal{C}_{\mathcal{S}'} \setminus \mathcal{C}^*$ is dominated in remote risk by some rule $c^* \in \mathcal{C}^*$.

⁶Consider for example the case where all agents have dogmatic priors but the weighted average prior is non-dogmatic.

Theorem 1. *The set of rules that are optimal in weighted average decision risk with respect to some weights $\omega \in \Omega$ is complete in decision risk. For nonempty $\mathcal{S}' \subseteq \mathcal{S}$, the set of rules that are optimal in weighted average remote risk with respect to some weights $\omega \in \Omega$ and signal space \mathcal{S}' is complete in remote risk with respect to \mathcal{S}' .*

For decision risk, Theorem 1 is an immediate consequence of classical complete class theorems for compact parameter spaces (see, e.g., Theorem 8.4.3 in Robert 2007). To prove the result for remote risk, we show that remote risk functions share the properties of decision risk functions used by the proof of the classical theorem. The public random variable V is important for this result, as it ensures that the class of remote risk functions is convex.

Theorem 1 states that any admissible rule (in either decision or remote risk) is weighted average optimal with respect to some weights. Limiting attention to full-support weights allows us to reverse the direction of the statement and show admissibility.

Lemma 6. *If a given rule $c \in \mathcal{C}_{\mathcal{D}}$ is optimal in weighted average decision risk with respect to weights $\omega \in \Omega_+$, then the rule c is admissible in decision risk. If a given rule $c \in \mathcal{C}_{\mathcal{S}'}$ is optimal in weighted average remote risk with respect to weights $\omega \in \Omega_+$ and nonempty $\mathcal{S}' \subseteq \mathcal{S}$, then the rule c is admissible in remote risk with respect to \mathcal{S}' .*

Together, Theorem 1 and Lemma 6 can be used to simplify the search for admissible rules. From Theorem 1, it is sufficient to search over the set of rules that are optimal in weighted average risk with respect to some weights. From Lemma 6, any rule that is optimal in weighted average risk with respect to full-support weights is admissible.

The close connection between weighted average risk optimality and admissibility implies that the admissibility conflicts encountered in the last section translate to conflicts for optimality in weighted average risk.

Corollary 1. *Suppose that there exists no rule $c \in \mathcal{C}_{\mathcal{D}}$ that is admissible in decision risk and also admissible in remote risk with respect to \mathcal{D} . Then any rule $\tilde{c} \in \mathcal{C}_{\mathcal{D}}$ that is optimal in weighted average decision risk with respect to weights $\tilde{\omega} \in \Omega_+$ is not optimal in weighted average remote risk with respect to any weights $\omega \in \Omega_+$ and signal space \mathcal{D} . Likewise, any rule $\tilde{c} \in \mathcal{C}_{\mathcal{D}}$ that is optimal in weighted average remote risk with respect to weights $\tilde{\omega} \in \Omega_+$ and signal space \mathcal{D} is not optimal in weighted average decision risk with respect to any weights $\omega \in \Omega_+$. This holds, in particular, if $N(\mathcal{X}, \mathcal{A}) \geq |\mathcal{D}|$ and there exists $d \in \mathcal{D}$ that is dominated in loss.*

Corollary 1 implies that in settings where there is an admissibility conflict between remote and decision risk (as in Examples 1 and 2), there exists no rule c and pair of weights $\omega_1, \omega_2 \in \Omega_+$ such that c is both optimal in weighted average decision risk with respect to ω_1 and optimal in weighted average remote risk with respect to ω_2 . Hence, for full-support weights ω not only is the Bayes decision rule with respect to the weighted average prior $a_\omega(\cdot)$ not optimal in weighted average remote risk with respect to ω , but also there exist no full-support weights that render this rule optimal in weighted average remote risk.

5 Maximum Risk

The final classical optimality criterion for decision rules that we study is minimaxity (see, e.g., Chapter 5 of Lehmann and Casella 1998). A minimax analysis avoids assigning weights by instead considering the worst-case risk.

Definition 9. *The **maximum decision risk** $\bar{R}(c)$ of a rule $c \in \mathcal{C}_{\mathcal{D}}$ is*

$$\bar{R}(c) = \max_{a \in \mathcal{A}} R_a(c).$$

The **maximum remote risk** $\bar{R}^*(c)$ of a rule $c \in \mathcal{C}$ is

$$\bar{R}^*(c) = \max_{a \in \mathcal{A}} R_a^*(c).$$

The proof of Theorem 1 shows that both decision risk and remote risk are continuous in a for all rules c , so since \mathcal{A} is compact these maxima are attained.

The maximum decision risk $\bar{R}(c)$ is the worst-case Bayes risk over a class of priors, and so is what Berger (1985) calls a Γ -Bayes risk. If (as in Examples 1 and 2) $\mathcal{A} = \Delta(\Theta)$, then $\bar{R}(c)$ is simply the maximum frequentist risk over Θ . The maximum remote risk $\bar{R}^*(c)$ is then the natural analogue of $\bar{R}(c)$ for remote risk. As with weighted average risk, we can rank procedures in terms of their maximum risk and use this to define optimal rules.

Definition 10. A rule $c \in \mathcal{C}_{\mathcal{D}}$ is **minimax in decision risk** if

$$\bar{R}(c) = \inf_{\tilde{c} \in \mathcal{C}_{\mathcal{D}}} \bar{R}(\tilde{c}).$$

For nonempty $\mathcal{S}' \subseteq \mathcal{S}$, a rule $c \in \mathcal{C}_{\mathcal{S}'}$ is **minimax in remote risk** with respect to \mathcal{S}' if

$$\bar{R}^*(c) = \inf_{\tilde{c} \in \mathcal{C}_{\mathcal{S}'}} \bar{R}^*(\tilde{c}).$$

We now ask whether we can always find a rule $c \in \mathcal{C}_{\mathcal{D}}$ that is minimax in decision risk and minimax in remote risk with respect to \mathcal{S} . Such a rule always exists if the set \mathcal{A} is convex.

Assumption 1. The set \mathcal{A} is convex.

If \mathcal{A} is convex then we can take arbitrary weighted averages of priors in \mathcal{A} without leaving the set. Intuitively, this means that for any two agents in the audience, we can always find a continuum of other agents in the audience with beliefs “between” these two. Assumption 1 holds when $\mathcal{A} = \Delta(\Theta)$, as in Examples 1 and 2.

Convexity allows us to apply a minimax theorem (from Grünwald and Dawid 2004, who in turn build on the classic minimax theorem of von Neumann 1928) to exchange the order of minimization and maximization in the definition of maximum remote risk. This greatly simplifies the problem of finding a minimax rule.

Lemma 7. *Under Assumption 1, for any $c \in \mathcal{C}^N = \mathcal{C}_S^N$,*

$$\bar{R}^*(c) = \max_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} R_a(b \circ c) = \min_{b \in \mathcal{B}} \max_{a \in \mathcal{A}} R_a(b \circ c) = \min_{b \in \mathcal{B}} \bar{R}(b \circ c).$$

Recall that \mathcal{C}^N is the class of non-randomized rules $c : \mathcal{X} \times [0, 1] \rightarrow \mathcal{S}$, while \mathcal{B} is the set of binding decision rules $b : \mathcal{S} \times [0, 1] \rightarrow \Delta(\mathcal{D})$. To minimize remote risk, we thus want to minimize $\min_{b \in \mathcal{B}} \bar{R}(b \circ c)$ over $c \in \mathcal{C}^N$. We can re-write this as a single minimization over \mathcal{C}_D .

Lemma 8. $\min_{c \in \mathcal{C}^N} \min_{b \in \mathcal{B}} \bar{R}(b \circ c) = \min_{c \in \mathcal{C}_D} \bar{R}(c)$.

Combining these results shows that the minimax decision risk and the minimax remote risk are the same. We further show that any rule that is minimax in decision risk is also minimax in remote risk with respect to the full signal space.

Theorem 2. *Under Assumption 1,*

$$\min_{c \in \mathcal{C}_D} \bar{R}(c) = \min_{c \in \mathcal{C}} \bar{R}^*(c).$$

Moreover, for any c^* with $\bar{R}(c^*) = \min_{c \in \mathcal{C}_D} \bar{R}(c)$ there exists $a^* \in \mathcal{A}$ with

$$\bar{R}(c^*) = R_{a^*}(c^*) = R_{a^*}^*(c^*) = \min_{c \in \mathcal{C}} R_{a^*}^*(c). \quad (5)$$

Consequently, $\bar{R}^*(c^*) = \min_{c \in \mathcal{C}} \bar{R}^*(c)$, so any rule that is minimax in decision risk is also minimax in remote risk with respect to \mathcal{S} .

Intuitively, we can think of a^* as an agent with a worst-case or least-favorable prior, for whom the minimax rule is an optimal decision rule. Theorem 2 shows

that, unlike for admissibility and weighted average risk, if we take minimaxity as our optimality criterion we can reconcile the recommendations of decision and remote risk.

Example 1 (Continued) Assumption 1 holds trivially in this example. Hence, Theorem 2 applies, and a rule that is minimax in decision risk is also minimax in remote risk. \blacktriangle

Example 2 (Continued) Assumption 1 again holds trivially in this example. In addition, there is a minimax decision rule c^* and corresponding worst-case prior $a^*(\cdot)$ with a simple form.

Claim 5. In Example 2, the rule $c^* \in \mathcal{C}_{\mathcal{D}}$ that takes $C^* = X$ whenever $X \in \Theta$ and takes C^* to randomize uniformly over Θ otherwise is minimax and admissible in decision risk. The rule c^* is likewise minimax in remote risk with respect to \mathcal{S} , but is dominated in remote risk by the rule c^{**} that takes $C^{**} = X$. Moreover, the agent a^* with the uniform prior $a^*(\theta) = 1/|\Theta|$ for all θ satisfies (5).

In the proof of Claim 5, it is crucial that the agent a^* is indifferent across all $d \in \Theta$ conditional on $X = \iota$, and so is happy with randomization in this case. This connects to the convexity of \mathcal{A} , which plays an important role in Theorem 2. To appreciate this point, let us consider a modified version of this example with $\mathcal{A} = \Delta(\Theta) \setminus \mathcal{N}_{\rho}(a^*)$ for $\mathcal{N}_{\rho}(a^*)$ the open ball of radius ρ around a^* . With this modification, \mathcal{A} is no longer convex, so Theorem 2 does not apply, and indeed the conclusion of the theorem does not hold for ρ small.

Claim 6. If we take $\mathcal{A} = \Delta(\Theta) \setminus \mathcal{N}_{\rho}(a^*)$, for ρ sufficiently small the rule c^* described in Claim 5 remains minimax in decision risk, but is no longer minimax in remote risk with respect to \mathcal{S} .

Intuitively, the conclusion of Theorem 2 fails in this case because once agent a^* is removed from the audience, all agents strictly prefer some action other than uniform

randomization conditional on $X = \iota$. Agents close to a^* still set $d = C^*$ under the rule c^* , however, and have the largest remote risk. Remote risk for these agents, and hence the maximum remote risk, is therefore strictly lower under the rule c^{**} . \blacktriangle

Remark 8. Inadmissibility of minimax rules. Theorem 2 does not allow us to avoid the conflicts between decision and remote risk optimality discussed in Sections 3 and 4. In particular, under the conditions of Proposition 2 the rule $c^* \in \mathcal{C}_{\mathcal{D}}$ defined in Theorem 2 is inadmissible in decision risk, inadmissible in remote risk with respect to \mathcal{D} (and thus also with respect to \mathcal{S}), or both. Indeed, Theorem 2 implies that (under Assumption 1) there is no improvement in worst-case risk from enlarging the signal space beyond \mathcal{D} , so when $N(\mathcal{X}, \mathcal{A}) > |\mathcal{D}|$, Lemma 3 implies that there are minimax rules that are dominated in remote risk.

6 Conclusions

We study the problem of communicating a summary of data to a diverse audience, and we contrast the conventional decision-theoretic notion of risk (which we term decision risk) with an alternative notion (which we term remote risk) that allows audience members to choose their own optimal decisions in light of the statistician's report. We show that there is sometimes no rule that is admissible in both notions of risk, and hence no rule that minimizes weighted average risk with respect to full-support weights under both notions of risk. By contrast, if the set of priors held by the audience is convex, rules that are minimax in decision risk are also minimax in remote risk.

Our findings suggest that a scientist interested in choosing rules based on admissibility or weighted average risk must, in general, decide whether she is concerned with recommending a good decision or with communicating the greatest amount of useful information. This choice is less stark for a scientist interested in choosing rules based on minimaxity. Even in this case, however, our results show that minimax rules can

“leave something on the table” from the perspective of either decision or remote risk (or both).

A possible way to resolve the tension between decision risk and remote risk is to convey two reports: a piece of information for those agents sophisticated enough to use it, and a recommended decision (estimator) for the rest. In general, the best report of the first kind will be the data X or the likelihood $F_\theta(X)$. When communication or information-processing constraints prevent conveying these rich objects, the scientist may look for summary information that is less rich but still very informative. This may explain the common practice of reporting “descriptive statistics” that are not directly interpretable as estimators but are nevertheless thought to convey useful information.

References

- Abadie, Alberto. 2018. Statistical non-significance in empirical economics. *MIT Working Paper*. Accessed on November 14 2018 at <https://economics.mit.edu/files/14851>.
- Abadie, Alberto and Maximilian Kasy. Forthcoming. Choosing among regularized estimators in empirical economics. *Review of Economics and Statistics*.
- Andrews, Isaiiah, Matthew Gentzkow, and Jesse M. Shapiro. 2017. Measuring the sensitivity of parameter estimates to estimation moments. *Quarterly Journal of Economics* 132(4): 1553-1592.
- Berger, James O. 1985. *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer-Verlag.
- Bergemann, Dirk and Stephen Morris. Forthcoming. Information design: A unified perspective. *Journal of Economic Literature*.
- Blackwell, David. 1951. The comparison of experiments. In J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*: 93-102. Berkeley: University of California Press.
- Blackwell, David. 1953. Equivalent comparisons of experiments. *Annals of Mathematical Statistics* 24(2): 265-272.

- Brown, Lawrence D. 1975. Estimation with incompletely specified loss functions (the case of several location parameters). *Journal of the American Statistical Association* 70(350): 417-427.
- Chamberlain, Gary and Edward E. Leamer. 1976. Matrix weighted averages and posterior bounds. *Journal of the Royal Statistical Society Series B (Methodological)* 38(1): 73-84.
- Cover, Thomas M. and Joy A. Thomas. 2006. *Elements of Information Theory*. 2nd ed. New York: Wiley-Interscience.
- Crawford, Vincent P. and Joel Sobel. 1982. Strategic information transmission. *Econometrica* 50(6): 1431-1451.
- Farrell, Joseph and Robert Gibbons. 1989. Cheap talk with two audiences. *American Economic Review* 79(5): 1214-1223.
- Frankel, Alex and Maximilian Kasy. 2018. Which findings should be published? *Harvard University Working Paper*. Accessed on November 8, 2018 at <<https://maxkasy.github.io/home/files/papers/findings.pdf>>.
- Geweke, John. 1997. Posterior simulators in econometrics. In D. M. Kreps and K. F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress* 3: 128-165. Cambridge: Cambridge University Press.
- Geweke, John. 1999. Using simulation methods for Bayesian econometric models: Inference, development, and communication. *Econometric Reviews* 18(1): 1-73.
- Giacomini, Raffaella and Toru Kitagawa. 2015. Robust inference about partially identified SVARs. *Cemmap Working Paper*.
- Grünwald, Peter D. and A. Philip Dawid. 2004. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics* 32(4): 1367-1433.
- Hildreth, Clifford. 1963. Bayesian statisticians and remote clients. *Econometrica* 31(3): 422-438.
- Kamenica, Emir and Matthew Gentzkow. 2011. Bayesian persuasion. *American Economic Review* 101(6): 2590-2615.
- Kwan, Yum K. 1999. Asymptotic Bayesian analysis based on a limited information estimator. *Journal of Econometrics* 88(1): 99-121.
- Leamer, Edward E. 1982. Sets of posterior means with bounded variance priors.

- Econometrica* 50(3): 725-736.
- Le Cam, L. 1996. Comparison of experiments—A short review. In T. S. Ferguson, L. S. Shapley, and J. B. MacQueen (eds.), *Statistics, Probability and Game Theory: Papers in Honor of David Blackwell*: 127-138. Hayward: Institute of Mathematical Statistics.
- Lehmann, E.L. and George Casella. 1998. *Theory of Point Estimation*. 2nd ed. New York: Springer-Verlag.
- Morris, Stephen. 1995. The common prior assumption in economic theory. *Economics and Philosophy* 11(2): 227-253.
- Müller, Ulrich K. and Andriy Norets. 2016. Coverage inducing priors in nonstandard inference problems. *Journal of the American Statistical Association* 111(515): 1233-1241.
- Poirier, Dale J. 1988. Frequentist and subjectivist perspectives on the problems of model building in economics. *Journal of Economic Perspectives* 2(1): 121-144.
- Pratt, John W. 1965. Bayesian interpretation of standard inference statements. *Journal of the Royal Statistical Society, Series B (Methodological)* 27(2): 169-203.
- Raiffa, Howard and Robert Schlaifer. 1961. *Applied Statistical Decision Theory*. Boston: Division of Research, Graduate School of Business Administration, Harvard University.
- Robert, Christian. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. 2nd ed. New York: Springer-Verlag.
- Sims, Christopher A. 1982. Scientific standards in econometric modeling. In M. Hazewinkel and A. H. G. Rinnooy Kan (eds.), *Current Developments in the Interface: Economics, Econometrics, Mathematics*: 317-337. Holland: D Reidel Publishing Company.
- Sims, Christopher A. 2007. Bayesian methods in applied econometrics, or, why econometrics should always and everywhere be Bayesian. Slides from the Hotelling Lecture, presented June 29, 2007 at Duke University.
- Torgersen, E. 1991. *Comparison of Experiments*. Encyclopedia of Mathematics and its Applications. 1st ed. Cambridge: Cambridge University Press.
- von Neumann, John. 1928. Zur theorie der gesellschaftsspiele. *Mathematische Annalen* 100(1): 295-320.

A Proofs

Proof of Lemma 1 Note that we can construct a $U [0, 1]$ random variable U independent of X and V such that for some \bar{c} , $C = \bar{c}(X, V, U)$ with probability one. For example, numbering the elements of \mathcal{S} as $s_1, \dots, s_{|\mathcal{S}|}$ and defining $c_j(X, V) = \Pr\{C = s_j | X, V\}$, conditional on $C = s_j$ let us take U to be uniformly distributed on

$$\left[\sum_{j' < j} c_{j'}(X, V), \sum_{j' \leq j} c_{j'}(X, V) \right].$$

Note that $U \sim U[0, 1]$ conditional on (X, V) for all (X, V) , and so is independent of (X, V) . If we then define $\bar{c}(X, V, U) = s_j$ if and only if

$$U \in \left[\sum_{j' < j} c_{j'}(X, V), \sum_{j' \leq j} c_{j'}(X, V) \right],$$

we see that $\bar{c}(X, V, U) = C$ with probability one, as desired.

Next, note that using V we can generate two independent $U [0, 1]$ random variables (V_1, V_2) , e.g. by taking alternating terms in the decimal expansion of V . Let us define $\tilde{c}_j(X, V) = 1\{\bar{c}(X, V_1, V_2) = s_j\}$ and $\tilde{C} = \bar{c}(X, V_1, V_2)$, noting that \tilde{C} is non-random conditional on (X, V) .

If we begin with $c \in \mathcal{C}_{\mathcal{D}}$, note that $\tilde{c} \in \mathcal{C}_{\mathcal{D}}$ as well, and that the conditional distribution of \tilde{C} conditional on X is the same as the conditional distribution of C given X for all X by construction. Hence, $\mathbb{E}_{\theta}[L(C, \theta)] = \mathbb{E}_{\theta}[L(\tilde{C}, \theta)]$ for all $\theta \in \Theta$ so $R_a(c) = R_a(\tilde{c})$ for all $a \in \mathcal{A}$ (where \mathbb{E}_{θ} denotes an expectation under fixed θ) which establishes the result for decision risk.

For remote risk, let us consider a rule $c \in \mathcal{C}_{\mathcal{S}'}$. For any $a \in \mathcal{A}$, let $b_a \in \mathcal{B}$ be such that

$$R_a(b_a \circ c) = \mathbb{E}_a[L(b_a(C, V), \theta)] = R_a^*(c)$$

noting that we can construct such a b_a by taking

$$b_a(C, V) \in \arg \min_{d \in \mathcal{D}} \mathbb{E}_a[L(d, \theta) | C, V].$$

Since $R_a(b_a \circ c) = R_a^*(c)$ by construction, and $b_a(\tilde{C}, V_1)$ is a feasible decision rule for agent a , we see that $R_a^*(\tilde{c}) \leq R_a^*(c)$.

To see why this inequality can be strict, it suffices to consider an example. In particular, suppose that $\mathcal{S} = \{s_1, s_2\}$, $\mathcal{X} = \{x_1, x_2\}$, and that c implies that $\Pr\{C = s_1|V, X = x_1\} = 2/3$ and $\Pr\{C = s_2|V, X = x_2\} = 2/3$. To achieve the same conditional distribution over $\tilde{C}|X$ with a non-randomized rule \tilde{c} , it must be the case that $\mathcal{V}_{11} = \{v|\tilde{c}(x_1, v) = s_1\}$ and $\mathcal{V}_{22} = \{v|\tilde{c}(x_2, v) = s_2\}$ both have Lebesgue measure $2/3$, and thus that $\mathcal{V}_{11} \cap \mathcal{V}_{22}$ has Lebesgue measure at least $1/3$ (since the measure of $\mathcal{V}_{11} \cup \mathcal{V}_{22}$ cannot exceed one). Conditional on $V \in \mathcal{V}_{11} \cap \mathcal{V}_{22}$, however, agents can perfectly infer X based on observing (\tilde{C}, V) , whereas this is never possible based on observing (C, V) . Hence, for certain loss functions and priors, remote risk will be strictly smaller under \tilde{c} than under c . \square

Proof of Proposition 1 For some $\tilde{a} \in \mathcal{A}$ and all $x \in \mathcal{X}$, let

$$c_{\tilde{a}}^*(x) \in \min_{d \in \mathcal{D}} E_{\tilde{a}}[L(d, \theta) | X = x]$$

select an optimal action with respect to $\tilde{a}'s$ posterior after x is observed. Note that $c_{\tilde{a}}^*$ is an optimal decision rule for \tilde{a} , in the sense that $R_{\tilde{a}}(c_{\tilde{a}}^*) = \min_{c \in \mathcal{C}_{\mathcal{D}}} R_{\tilde{a}}(c)$. Moreover, since $c_{\tilde{a}}^*$ is the action \tilde{a} would take if they could observe the full data,

$$R_{\tilde{a}}(c_{\tilde{a}}^*) = R_{\tilde{a}}^*(c_{\tilde{a}}^*) = \min_{c \in \mathcal{C}} R_{\tilde{a}}^*(c).$$

Hence, it is impossible to ε -dominate $c_{\tilde{a}}^*$ in either decision risk or remote risk for any ε , and we have constructed a rule that is extended-admissible in both decision risk and remote risk. \square

Proof of Lemma 2 Follows as a consequence of Lemmas 5 and 6 which are proved below. \square

Proof of Lemma 3 We first show that if $|\mathcal{S}''| \geq N(\mathcal{X}, \mathcal{A})$ then there exists a rule $c^* \in \mathcal{C}_{\mathcal{S}''}$ that is admissible in remote risk with respect to \mathcal{S}' . Since $|\mathcal{S}''| \geq N(\mathcal{X}, \mathcal{A})$, we know that there exists a partition $P \in \mathcal{P}^*$ with $|P| \leq |\mathcal{S}''|$, for \mathcal{P}^* as in Definition 4. Let c^* be a rule that associates a unique signal with each cell of P , so for $P = \{\mathcal{X}_1, \dots, \mathcal{X}_{|P|}\}$ and $\mathcal{S}'' = \{s_1, \dots, s_{|\mathcal{S}''|}\}$, $C^* = s_j$ if and only if $X \in \mathcal{X}_j$. By the

definition of \mathcal{P}^* , for each $a \in \mathcal{A}$ and each $j \in \{1, \dots, |P|\}$, there exists $d_{a,j}$ such that

$$d_{a,j} \in \arg \min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | X] \text{ for all } X \in \mathcal{X}_j.$$

If agent a adopts the rule that after observing C^* they take the decision $d_a(C^*)$, where $d_a(C^*) = d_{a,j}$ if and only if $C^* = s_j$, they therefore obtain the same risk as if they took their optimal action given X . Since no rule in $\mathcal{C}_{\mathcal{S}'}$ can achieve strictly lower risk for agent a , and this argument holds for all $a \in \mathcal{A}$, this shows that c^* is admissible in remote risk with respect to \mathcal{S}' .

We next show that if $|\mathcal{S}''| < N(\mathcal{X}, \mathcal{A})$, then any rule $c \in \mathcal{C}_{\mathcal{S}''}$ is dominated by some rule in $\mathcal{C}_{\mathcal{S}'}$. By Lemma 1 we can limit attention to rules c that use only the public randomization device, and hence where C is non-random conditional on X and V . In a minor abuse of notation, we will write $C = c(X, V)$ in this case. Since $|\mathcal{S}''| < N(\mathcal{X}, \mathcal{A})$, we know that for every V there is some $a_V \in \mathcal{A}$, some $\mathcal{X}_V \subseteq \mathcal{X}$, and some $s_V \in \mathcal{S}$ such that

$$\bigcap_{X \in \mathcal{X}_V} \arg \min_{d \in \mathcal{D}} \mathbb{E}_{a_V} [L(d, \theta) | X] = \emptyset$$

and $c(X, V) = s_V$ for all $X \in \mathcal{X}_V$.⁷

We can view (\mathcal{X}_V, s_V) as a random variable supported on $2^{\mathcal{X}} \times \mathcal{S}$, for $2^{\mathcal{X}}$ the power set of \mathcal{X} . Since \mathcal{X}_V is non-empty with probability one and $(\mathcal{X}, \mathcal{S})$ are finite, there exists a pair $\mathcal{X}^* \subseteq \mathcal{X}$ and $s^* \in \mathcal{S}$ such that \mathcal{X}^* is non-empty and $\Pr \{\mathcal{X}_V = \mathcal{X}^*, s_V = s^*\} > 0$, where this event depends only on V so the probability does not need to be subscripted by a . Let a^* be an agent with

$$\bigcap_{X \in \mathcal{X}^*} \arg \min_{d \in \mathcal{D}} \mathbb{E}_{a^*} [L(d, \theta) | X] = \emptyset, \tag{6}$$

noting that we can take $a^* = a_V$ for any V with $\mathcal{X}_V = \mathcal{X}^*$. Let

$$\mathcal{V}^* = \{V \in [0, 1] : \mathcal{X}_V = \mathcal{X}^*, s_V = s^*\},$$

⁷If this is not the case, then the level sets of $c(\cdot, V)$ form a partition in \mathcal{P}^* of size at most $|\mathcal{S}''|$, contradicting the assumption that $|\mathcal{S}''| < N(\mathcal{X}, \mathcal{A})$.

so $s^* = C(X, V)$ for all $X \in \mathcal{X}^*$ and $V \in \mathcal{V}^*$. Finally, let d^* denote an element of

$$\arg \min_{d \in \mathcal{D}} \mathbb{E}_{a^*} [L(d, \theta) | V \in \mathcal{V}^*, C = s^*].$$

By (6), there exists $\tilde{X} \in \mathcal{X}^*$ and $\tilde{d} \in \mathcal{D}$ such that

$$\mathbb{E}_{a^*} [L(\tilde{d}, \theta) | \tilde{X}] < \mathbb{E}_{a^*} [L(d^*, \theta) | \tilde{X}].$$

For \tilde{s} an element of $\mathcal{S}' \setminus \mathcal{S}''$, let \tilde{c} be the rule that reports $\tilde{C} = \tilde{s}$ when $V \in \mathcal{V}^*$ and $X = \tilde{X}$, and that agrees with c otherwise. Given \tilde{C} an agent can reconstruct C , so the rule \tilde{c} does not increase remote risk for any agent relative to C . For agent a^* , however,

$$\mathbb{E}_{a^*} [L(\tilde{d}, \theta) | \tilde{C} = \tilde{s}, V \in \mathcal{V}^*] < \mathbb{E}_{a^*} [L(d^*, \theta) | \tilde{C} = \tilde{s}, V \in \mathcal{V}^*],$$

so d^* is now sub-optimal for a^* conditional on observing $\{\tilde{C} = \tilde{s}, V \in \mathcal{V}^*\}$. This immediately implies that

$$\mathbb{E}_{a^*} \left[\min_{d \in \mathcal{D}} \mathbb{E}_{a^*} [L(d, \theta) | \tilde{C}, V \in \mathcal{V}^*] | C = s^*, V \in \mathcal{V}^* \right] < \min_{d \in \mathcal{D}} \mathbb{E}_{a^*} [L(d, \theta) | C = s^*, V \in \mathcal{V}^*].$$

Since $\Pr_{a^*} \{C = s^*, V \in \mathcal{V}^*\} > 0$, this implies that $R_{a^*}^*(\tilde{c}) < R_{a^*}^*(c)$, so \tilde{c} dominates c as we wanted to show. \square

Proof of Claim 1 To prove the claim, we will show that for any pair $X, X' \in \mathcal{X}$ with $X \neq X'$, there exists an agent $a \in \mathcal{A}$ such that

$$\arg \min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | X] \cap \arg \min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | X'] = \emptyset, \quad (7)$$

since this implies that the only element of \mathcal{P}^* is the trivial partition.

Note that by the monotone likelihood ratio assumption the posterior distribution for any agent with a non-degenerate prior is strictly increasing in X in the sense of first-order stochastic dominance. Note, next, that the assumption that $\mathcal{A} = \Delta(\Theta)$ along with the assumption of full support for F_θ for all $\theta \in \Theta$ implies that the class of posteriors conditional on X remains equal to $\Delta(\Theta)$ for all X . The posterior risk

function for agent a conditional on X is

$$\sum_{\theta \in \Theta} l(|d - \theta|) \pi_a(\theta|X).$$

Strict convexity implies that each agent's set of optimal actions conditional on X contains at most two elements. Let us fix a value $X < \max\{\mathcal{X}\}$ and consider an agent a who, conditional on X , is indifferent between two decisions d and d' with $d < d'$. Note that since agent a is indifferent between d and d' , it must be that for some $\theta_L \leq d$ and $\theta_H \geq d'$, $\Pr_a\{\theta = \theta_L\} \in (0, 1)$ and $\Pr_a\{\theta = \theta_H\} \in (0, 1)$. If we define a_ε as the agent with $\varepsilon > 0$ more prior mass on θ_L and ε less prior mass on θ_H relative to a , then for ε sufficiently small a_ε strictly prefers d to any other action conditional on X , but strictly prefers some $d'' \geq d'$ conditional on any $X' > X$ (where our assumptions imply that $|\mathcal{D}| \geq 3$). Since we can repeat this argument for all $X < \max\{\mathcal{X}\}$, this verifies (7) and so proves the claim. \square

Proof of Claim 2 To prove the claim, we again need to show that (7) holds for all pairs $X, X' \in \mathcal{X}$ with $X \neq X'$. Let us define \mathcal{A}^* as the subset of distributions in $\Delta(\Theta)$ with $a(\theta')/a(\theta'') < \frac{1-\delta-\gamma}{\delta/(|\Theta|-1)}$ for all $\theta', \theta'' \in \Theta$. Note that for any $a \in \mathcal{A}^*$ and any $X \in \Theta$, a 's posterior odds satisfy

$$\frac{a(\theta = x|X = x)}{a(\theta = \theta'|X = x)} = \frac{a(\theta = x)}{a(\theta = \theta')} \frac{1 - \delta - \gamma}{\delta / (|\Theta| - 1)} > 1$$

for all $\theta' \neq x$. Hence, conditional on $X = x \in \Theta$, x is the unique posterior mode (and hence the unique optimal decision) for all agents $a \in \mathcal{A}^*$. Note, next, that conditional on $X = \iota$ the posterior mode for each agent a is simply their prior mode, and that the set of prior modes for agents in \mathcal{A}^* is Θ . Thus, (7) holds. \square

Proof of Lemma 4 By Definition 5, there exists $d' \in \mathcal{D}$ such that $L(d, \theta) \geq L(d', \theta)$ for all θ , with strict inequality for some θ' such that $\Pr_{a'}\{\theta = \theta'\} > 0$ for some $a' \in \mathcal{A}$. Consider any rule c that chooses d with positive probability conditional on some X . Note that if we consider the alternative rule which chooses d' instead, we weakly reduce risk for all agents. Moreover, we strictly reduce risk for agent a' since this agent assigns positive prior probability to θ' and X has full support conditional on θ' . \square

Proof of Proposition 2 From Lemma 3, since $|\mathcal{D}| \leq N(\mathcal{X}, \mathcal{A})$, any rule $c \in \mathcal{C}_{\mathcal{D}'}$ for $\mathcal{D}' = \mathcal{D}/\{d\}$ is inadmissible in remote risk with respect to \mathcal{D} . But by Lemma 4, any rule $c \in \mathcal{C}_{\mathcal{D}}/\mathcal{C}_{\mathcal{D}'}$ is inadmissible in decision risk. Therefore, no rule $c \in \mathcal{C}_{\mathcal{D}}$ is both admissible in remote risk with respect to \mathcal{D} and admissible in decision risk. \square

Proof of Claim 3 For any $d \in \mathcal{D}$ with $d > d' = \max\{\Theta\}$, strict convexity of $l(\cdot)$ implies that $L(d, \theta) > L(d', \theta)$ for all $\theta \in \Theta$. Hence, any such d is dominated in loss. \square

Proof of Claim 4 For $d = \iota$ and any $d' \in \Theta$, $L(d, \theta) \geq L(d', \theta)$, with a strict inequality at $\theta = d'$. Hence, $d = \iota$ is dominated in this loss. \square

Proof of Lemma 5 We first show the result for remote risk. Let us consider rules in $\mathcal{C}_{\mathcal{S}'}$ for some non-empty $\mathcal{S}' \subseteq \mathcal{S}$. We prove the result by proving it for each V . In particular, we want to show that there exists a function $c^* \in \mathcal{C}_{\mathcal{S}'}$ such that

$$\int \mathbb{E}_a [\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C^*, V] | V] d\omega(a) = \inf_{c \in \mathcal{C}_{\mathcal{S}'}} \int \mathbb{E}_a [\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V] d\omega(a) \quad (8)$$

for all V .

To prove this result, note that for fixed V we can view c as a function from \mathcal{X} to $\Delta(\mathcal{S}')$, $c: \mathcal{X} \rightarrow \Delta(\mathcal{S}')$. Hence, for fixed V we can view each rule c as an element of $\Delta(\mathcal{S}')^{|\mathcal{X}|}$, and the set of rules is compact under the usual metric. To guarantee that (8) holds, it thus suffices to show that $\int \mathbb{E}_a [\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V] d\omega(a)$ is continuous as a function of c .

Note, however, that since \mathcal{D} and Θ are finite, the maximum loss is bounded, so for any $c, \tilde{c} \in \mathcal{C}_{\mathcal{S}'}$ and any $b \in \mathcal{B}$,

$$\begin{aligned} & \sup_{b \in \mathcal{B}} \left| \mathbb{E}_a [L(b(C, V), \theta) | V] - \mathbb{E}_a [L(b(\tilde{C}, V), \theta) | V] \right| \\ & \leq 2 \max_{d, \theta} |L(d, \theta)| \sum_{(x, s) \in \mathcal{X} \times \mathcal{S}} \left| \Pr_a \{X = x, C = s | V\} - \Pr_a \{X = x, \tilde{C} = s | V\} \right|. \end{aligned}$$

Assume without loss of generality that

$$\mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V \right] \geq \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | \tilde{C}, V] | V \right],$$

and note that for $b_a^{\tilde{c}}(\tilde{C}, V) \in \arg \min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | \tilde{C}, V]$ for all C ,

$$\mathbb{E}_a [\mathbb{E}_a [L(b_a^{\tilde{c}}(C, V), \theta) | C, V] | V] \geq \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V \right].$$

This implies that

$$\begin{aligned} & \left| \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V \right] - \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | \tilde{C}, V] | V \right] \right| \\ &= \left| \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V \right] - \mathbb{E}_a [L(b_a^{\tilde{c}}(\tilde{C}, V), \theta) | V] \right| \\ &\leq \left| \mathbb{E}_a [L(b_a^{\tilde{c}}(C, V), \theta) | V] - \mathbb{E}_a [L(b_a^{\tilde{c}}(\tilde{C}, V), \theta) | V] \right| \\ &\leq 2 \max_{d, \theta} |L(d, \theta)| \sum_{(x, s) \in \mathcal{X} \times \mathcal{S}} \left| \Pr_a \{X = x, C = s | V\} - \Pr_a \{X = x, \tilde{C} = s | V\} \right| \end{aligned}$$

Hence,

$$\begin{aligned} & \left| \int \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V \right] d\omega(a) - \int \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | \tilde{C}, V] | V \right] d\omega(a) \right| \\ &\leq \int \left| \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V \right] - \mathbb{E}_a \left[\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | \tilde{C}, V] | V \right] \right| d\omega(a) \\ &\leq 2 \max_{d, \theta} |L(d, \theta)| \int \sum_{(x, s) \in \mathcal{X} \times \mathcal{S}} \left| \Pr_a \{X = x, C = s | V\} - \Pr_a \{X = x, \tilde{C} = s | V\} \right| d\omega(a) \end{aligned}$$

which tends to zero as $\|c - \tilde{c}\| \rightarrow 0$. Hence $\int \mathbb{E}_a [\min_{d \in \mathcal{D}} \mathbb{E}_a [L(d, \theta) | C, V] | V] d\omega(a)$ is continuous in c , and the infimum is achieved. Since this argument applies for each V , it follows that the $\inf_{c \in \mathcal{C}_S} R^*(c)$ is achieved as well. The same argument, omitting minimization over \mathcal{B} , establishes the result for decision risk. \square

Proof of Theorem 1 Since \mathcal{A} is a closed subset of the simplex, it is compact.

We first prove the result for decision risk. Note that the set of decision risk functions is trivially convex, since we can take a mixture of any two decision rules to form a new decision rule. Hence we can apply Theorem 8.4.3 of Robert (2007) (with

\mathcal{A} playing the role of Θ) to obtain that weighted average risk minimizing procedures are a complete class for decision risk.

We next prove the result for remote risk. Note first that the presence of V ensures that the set of remote risk functions is convex. Specifically, for any pair of rules $c_1, c_2 \in \mathcal{C}_{S'}$ and any $\alpha \in [0, 1]$ we can construct a new rule

$$c_\alpha(X, V) = 1 \{V \leq \alpha\} c(X, V/\alpha) + 1 \{V > \alpha\} c(X, (1 - V)/(1 - \alpha)).$$

Since each agent a observes (C_α, V) ,

$$R_a^*(c_\alpha) = \alpha R_a^*(c_1) + (1 - \alpha) R_a^*(c_2).$$

We show in the proof of Lemma 6 that remote risk is continuous in a , so combined with the convexity of the class of communication rules, Theorem 8.4.3 of Robert (2007) again implies that weighted average risk minimizing procedures are a complete class for remote risk. \square

Proof of Lemma 6 We prove the result first for decision risk, and then for remote risk. Note, first, that the decision risk of a rule c can be written as

$$\sum_{\Theta} a(\theta) E_\theta [L(c(X, V), \theta)]$$

where $L(d, \theta)$ is uniformly bounded. Hence, the decision risk is trivially continuous in a . To prove admissibility of weighted average-optimal rules in decision risk, suppose that the result fails, so there exists some $c', c'' \in \mathcal{C}_{\mathcal{D}}$ where

$$\int R_a(c') d\omega(a) = \int R_a(c'') d\omega(a) = \inf_{c \in \mathcal{C}_{\mathcal{D}}} \int R_a(c) d\omega(a)$$

for $\omega \in \Omega_+$ but c'' dominates c' . This implies that there exists some $a \in \mathcal{A}$ with $R_a(c') > R_a(c'')$ and, by continuity of $R_a(c)$, that the same holds on an open neighborhood of a in $\Delta(\Theta)$ (which may or may not be contained in \mathcal{A}). The full support assumption on ω implies that the intersection of this open neighborhood with \mathcal{A} has

positive ω measure,⁸ and thus that

$$\int R_a(c') d\omega(a) < \int R_a(c'') d\omega(a).$$

Hence, we have achieved a contradiction and proved the result for decision risk.

For remote risk, we first show that the remote risk $R_a^*(c)$ is continuous in a for all c . To this end, note that by the same argument used to establish continuity of decision risk, for any fixed $c \in \mathcal{C}_{S'}$ and $b \in \mathcal{B}$, $E_a[L(b(C, V), \theta)]$ is continuous in a . Indeed,

$$\begin{aligned} \sup_{c,b} |E_a[L(b(C, V), \theta)] - E_{\tilde{a}}[L(b(C, V), \theta)]| \\ \leq 2 \max_{d,\theta} |L(d, \theta)| \sum_{\Theta} |a(\theta) - \tilde{a}(\theta)| \end{aligned} \quad (9)$$

so $E_a[L(b(C, V), \theta)]$ is Lipschitz in a uniformly over b, c .

Towards contradiction, let us assume that there exists c such that $R_a^*(c)$ is discontinuous in a at a^* . Then there exists a sequence of values $a_k \rightarrow a^*$ such that either

$$\limsup_{k \rightarrow \infty} R_{a_k}^*(c) < R_{a^*}^*(c)$$

or⁹

$$\liminf_{k \rightarrow \infty} R_{a_k}^*(c) > R_{a^*}^*(c).$$

First, consider the case with $\limsup_{k \rightarrow \infty} R_{a_k}^*(c) < R_{a^*}^*(c)$. For k sufficiently large and b_k such that

$$E_{a_k}[L(b_k(C, V), \theta)] = \min_{b \in \mathcal{B}} E_{a_k}[L(b(C, V), \theta)],$$

(9) implies that

$$\begin{aligned} E_{a^*}[L(b_k(C, V), \theta)] \\ \leq E_{a_k}[L(b_k(C, V), \theta)] + 2 \max_{d,\theta} |L(d, \theta)| \sum_{\Theta} |a^*(\theta) - a_k(\theta)| < R_{a^*}^*(c) \end{aligned}$$

contradicting the definition of remote risk. Likewise, if $\liminf_{k \rightarrow \infty} R_{a_k}^*(c) > R_{a^*}^*(c)$

⁸Since a is in the support of ω by assumption, we know that ω assigns positive mass to all open neighborhoods of a .

⁹Note that if $\limsup_{k \rightarrow \infty} R_{a_k}^*(c) \geq R_{a^*}^*(c)$ and $\liminf_{k \rightarrow \infty} R_{a_k}^*(c) \leq R_{a^*}^*(c)$ for all $a_k \rightarrow a^*$, then $R_{a^*}^*(c)$ is continuous at a^* .

then for b^* such that

$$\mathbb{E}_{a^*} [L(b^*(C, V), \theta)] = \min_{b \in \mathcal{B}} \mathbb{E}_{a^*} [L(b(C, V), \theta)],$$

and k sufficiently large we have

$$\mathbb{E}_{a_k} [L(b^*(C, V), \theta)] \leq \mathbb{E}_{a^*} [L(b^*(C, V), \theta)] + 2 \max_{d, \theta} |L(d, \theta)| \sum_{\Theta} |a^*(\theta) - a_k(\theta)| < R_{a_k}^*(c)$$

which is again a contradiction. Hence, remote risk is continuous in a , as we wanted to show. Given this continuity, the result for remote risk follows by the same argument used above for decision risk. \square

Proof of Corollary 1 Follows from Lemma 6 and Proposition 2. \square

Proof of Lemma 7 We prove this result using Theorems 5.1 and 5.2 of Grunwald and Dawid (2004). Let us fix a non-randomized rule $c \in \mathcal{C}$, and for $s \in \mathcal{S}$ define $\mathcal{V}(X, s) = \{v : c(X, v) = s\}$. Let us denote the elements of \mathcal{X} by $\{x_1, \dots, x_{|\mathcal{X}|}\}$, and let $s^{|\mathcal{X}|}$ denote a vector in $\mathcal{S}^{|\mathcal{X}|}$ with j th element $s_j^{|\mathcal{X}|} \in \mathcal{S}$. Let

$$\mathcal{V}^* = \left\{ \left\{ \bigcap_{j=1}^{|\mathcal{X}|} \mathcal{V}(x_j, s_j^{|\mathcal{X}|}) \right\} : s^{|\mathcal{X}|} \in \mathcal{S}^{|\mathcal{X}|} \right\},$$

and note that if we know that $V \in \tilde{\mathcal{V}}$ for a given (and known) $\tilde{\mathcal{V}} \in \mathcal{V}^*$, we know the value of $c(x, V)$ for all $x \in \mathcal{X}$. Hence, if we number the elements of \mathcal{V}^* as $\{\tilde{\mathcal{V}}_1, \dots, \tilde{\mathcal{V}}_{|\mathcal{V}^*|}\}$ and define a new (discrete) random variable $V^* = \sum_{j=1}^{|\mathcal{V}^*|} j 1_{\{V \in \tilde{\mathcal{V}}_j\}}$, we have $c(X, V) = \tilde{c}(X, V^*)$ for some function \tilde{c} , so C depends on X and V^* only. Moreover, V is independent of X conditional on (C, V^*) , so observing (C, V^*) is as informative for all agents as observing (C, V) . In particular, for $\tilde{b} : \mathcal{S} \times \{1, \dots, |\mathcal{V}^*|\} \rightarrow \Delta(\mathcal{D})$ a binding decision rule based on (C, V^*) and $\tilde{\mathcal{B}}$ the class of such rules, for any $b \in \mathcal{B}$ there exists $\tilde{b} \in \tilde{\mathcal{B}}$ such that $R_a(b \circ c) = R_a(\tilde{b} \circ \tilde{c})$ for all $a \in \mathcal{A}$ (and the reverse clearly holds as well).

Based on this fact, we can cast our problem into the setting of Grunwald and Dawid (2004)'s Theorems 5.1 and 5.2. Specifically, let us view $\tilde{b} \in \tilde{\mathcal{B}}$ as the action “ a ” in their terminology, and let us define the state “ X ” in their terminology as the triple (C, V^*, θ) , noting that this state takes only a finite number of values in our setting. Note, next, that since the set of priors \mathcal{A} is closed, and is convex by Assumption 1,

the set of implied distributions for (C, V^*, θ) is closed and convex as well. Finally, we note that the set of risk functions for $\tilde{b} : \mathcal{C} \times \{1, \dots, |\mathcal{V}^*|\} \rightarrow \mathcal{D}$ is finite, and hence trivially compact. The set of risk functions for (potentially randomized) binding decision rules $\tilde{b} \in \tilde{\mathcal{B}}$ is thus compact as well, which ensures that all the minima over \mathcal{B} in the statement of the Lemma are achieved. Hence, the assumptions of Theorems 5.1 and 5.2 of Grunwald and Dawid (2004) hold, and the result follows immediately. \square

Proof of Lemma 8 Note first that the set of decision risk functions $R_a(c)$ for decision rules $c : \mathcal{X} \rightarrow \mathcal{D}$ is finite and hence trivially compact, while the set of risk functions for $c \in \mathcal{C}_D^N$ or $c \in \mathcal{C}_D$ is the convex hull of a compact set, and so is compact as well. Hence, all the minima over \mathcal{C}_D^N and \mathcal{C}_D in the statement of the lemma are achieved.

Given any pair $c \in \mathcal{C}$ and $b \in \mathcal{B}$, if we define $\tilde{c} \in \mathcal{C}_D$ as $\tilde{c} = b \circ c$, we see that for every $x \in \mathcal{X}$ we induce the same distribution over \mathcal{D} , and thus that \tilde{c} and $b \circ c$ have the same decision risk. Hence, since $\mathcal{C}_D^N \subset \mathcal{C}$,

$$\min_{c \in \mathcal{C}_D^N} \min_{b \in \mathcal{B}} \max_{a \in \mathcal{A}} R_a(b \circ c) \geq \min_{c \in \mathcal{C}} \min_{b \in \mathcal{B}} \max_{a \in \mathcal{A}} R_a(b \circ c) \geq \min_{c \in \mathcal{C}_D} \max_{a \in \mathcal{A}} R_a(c).$$

Conversely, for any $\tilde{c} \in \mathcal{C}_D$, since $\mathcal{D} \subseteq \mathcal{S}$ we can take $c^*(x) = \tilde{c}(x)$ for all $x \in \mathcal{X}$, $b(\cdot) = b^I(\cdot)$ to be the identity mapping, and again obtain the same risk. Hence,

$$\min_{c \in \mathcal{C}} \min_{b \in \mathcal{B}} \max_{a \in \mathcal{A}} R_a(b \circ c) = \min_{c \in \mathcal{C}_D} \max_{a \in \mathcal{A}} R_a(c),$$

and we know that all minima are achieved. By the proof of Lemma 1, however, we further know that for any $c \in \mathcal{C}_D$ there exists a non-randomized $c' \in \mathcal{C}_D^N$ that yields the same distribution over \mathcal{S} , and thus that

$$\min_{c \in \mathcal{C}_D^N} \min_{b \in \mathcal{B}} \max_{a \in \mathcal{A}} R_a(b \circ c) = \min_{c \in \mathcal{C}} \min_{b \in \mathcal{B}} \max_{a \in \mathcal{A}} R_a(b \circ c),$$

which completes the proof. \square

Proof of Theorem 2 To prove the first part of the theorem, note that by Lemma 1,

$$\min_{c \in \mathcal{C}} \bar{R}^*(c) = \min_{c \in \mathcal{C}_D^N} \bar{R}^*(c).$$

By Lemma 7, however, for each $c \in \mathcal{C}^N$, $\bar{R}^*(c) = \min_{b \in \mathcal{B}} \bar{R}(b \circ c)$, so

$$\min_{c \in \mathcal{C}} \bar{R}^*(c) = \min_{c \in \mathcal{C}^N} \min_{b \in \mathcal{B}} \bar{R}(b \circ c).$$

Lemma 8 then implies that

$$\min_{c \in \mathcal{C}^N} \min_{b \in \mathcal{B}} \bar{R}(b \circ c) = \min_{c \in \mathcal{C}_D} \bar{R}(c),$$

which completes the proof of the first part of the theorem.

For the second part of the theorem, note that we already argued in the proof of Lemma 8 that there exists $c^* \in \mathcal{C}_D$ with $\min_{c \in \mathcal{C}_D} \bar{R}(c) = \bar{R}(c^*)$. Let $\tilde{c}^* : \mathcal{X} \rightarrow \Delta(\mathcal{D})$ be a rule that does not use the public randomization device and induces the same distribution for $\tilde{C}^*|X$ as that of $C^*|X$, and note that $R_a(\tilde{c}^*) = R_a(c^*)$ for all a , so $\bar{R}(c^*) = \bar{R}(\tilde{c}^*)$. For $\tilde{\mathcal{C}}_D$ the class of such rules, note that if we limit attention to $c \in \tilde{\mathcal{C}}_D$ we can cast the problem into the setting of Section 5 of Grunwald and Dawid (2004). Specifically, let us view $c \in \tilde{\mathcal{C}}_D$ as the action “ a ” in their terminology, and let us define the state “ X ” in their terminology as the pair (X, θ) , noting that this state takes only a finite number of values. Since the class of priors \mathcal{A} is closed and convex, the set of implied distributions for (X, θ) is closed and convex as well. Finally, the set of risk functions for $c : \mathcal{X} \rightarrow \mathcal{D}$ is finite, while the set of risk functions for $\tilde{\mathcal{C}}_D$ is the convex hull of a finite set and so is compact.

Theorem 5.2 of Grunwald and Dawid (2004) then implies that there exists $a^* \in \mathcal{A}$ such that

$$\bar{R}(\tilde{c}^*) = R_{a^*}(\tilde{c}^*) = \min_{c \in \tilde{\mathcal{C}}_D} R_{a^*}(c). \quad (10)$$

Note, however, that the the same argument used to construct \tilde{c}^* ,

$$\min_{c \in \tilde{\mathcal{C}}_D} R_{a^*}(c) = \min_{c \in \mathcal{C}_D} R_{a^*}(c),$$

so (10) implies that

$$\bar{R}(c^*) = R_{a^*}(c^*) = \min_{c \in \mathcal{C}_D} R_{a^*}(c).$$

Note, next, that by the definition of remote risk,

$$\min_{c \in \mathcal{C}_D} R_{a^*}(c) = \min_{c \in \mathcal{C}} R_{a^*}^*(c),$$

and thus that $R_{a^*}(c^*) = R_{a^*}^*(c^*)$. This establishes that

$$\bar{R}(c^*) = R_{a^*}(c^*) = R_{a^*}^*(c^*) = \min_{c \in \mathcal{C}} R_{a^*}^*(c),$$

as we aimed to show. Note, moreover, that by definition $\bar{R}^*(\tilde{c}) \geq \min_{c \in \mathcal{C}} R_{a^*}^*(c)$ for all $\tilde{c} \in \mathcal{C}_{\mathcal{D}}$, and that $R_a^*(c^*) \leq \bar{R}(c^*) = R_{a^*}(c^*)$ for all a , so $\min_{c \in \mathcal{C}} \bar{R}^*(c) = \bar{R}^*(c^*)$.

□

Proof of Claim 5 To prove minimaxity, note that for all c ,

$$\bar{R}(c) = \max_{\theta \in \Theta} R_{\theta}(c),$$

for $R_{\theta}(c)$ the decision risk under the prior that puts probability one on θ (or, equivalently, frequentist risk under θ). Hence, c is minimax in decision risk over \mathcal{A} if and only if it is minimax over Θ . Note, next, that c^* is invariant under permutation of the elements of Θ (see Chapter 3 of Lehmann and Casella 1998). The unique prior invariant under permutation is the uniform prior over Θ , $a^*(\theta) = 1/|\Theta|$ for all θ , and c^* is an optimal (Bayes) decision rule under this prior. Since the permutation group is transitive, Bayes decision rules under the invariant prior are minimax by Theorem 3.1 in Chapter 5 of Lehmann and Casella (1998), and so are admissible by that theorem as well (or by Lemma 6 above).

That c^* is minimax in remote risk follows from Theorem 2. To see that it is inadmissible in remote risk, we argue that it is dominated by the rule c^{**} that takes $C^{**} = X$. Note first that $R_a^*(c^{**}) \leq R_a^*(c^*)$ for all a , since based on observing C^{**} an agent can generate a random variable with the same distribution as C^* by randomizing over Θ whenever $C^{**} = \iota$. Finally, note that in the proof of Claim 6 below, we show that there are agents $a \in \mathcal{A}$ for whom $R_a^*(c^{**}) < R_a^*(c^*)$, which shows that c^* is dominated.

To verify that a^* satisfies (5), note that since c^* is an optimal decision rule for a^* , $R_{a^*}(c^*) = R_{a^*}^*(c^*)$. Note, moreover, that c^* is invariant under permutation of the elements of Θ , and so has a constant risk function over θ , and thus over \mathcal{A} . Hence, $R_a(c^*)$ is constant in a , so $\bar{R}(c^*) = R_{a^*}(c^*)$. Finally, note that by the definition of remote risk $R_a^*(c^*) \leq R_a(c^*)$ for all a , which together preceding argument implies that $\bar{R}^*(c^*) = R_{a^*}^*(c^*)$. □

Proof of Claim 6 Note, first, that provided $\rho < \frac{1}{|\Theta|}$ it remains the case that

$$\sup_{a \in \Delta(\Theta) \setminus \mathcal{N}_\rho(a^*)} R_a(c) = \max_{\theta \in \Theta} R_\theta(c),$$

so the argument for minimaxity of c^* in decision risk in the proof of Claim 5 continues to hold.

To complete the proof, we need to show that for ρ sufficiently small, c^* is not minimax in remote risk. To this end, recall that as shown in the proof of Claim 5, all agents have the same decision risk for c^* , $R_a(c^*) = \bar{R}(c^*)$ for all $a \in \Delta(\Theta) \setminus \mathcal{N}_\rho(a^*)$. Since remote risk is bounded above by decision risk, this implies that $R_a^*(c^*) \leq \bar{R}(c^*)$ for all $a \in \Delta(\Theta) \setminus \mathcal{N}_\rho(a^*)$.

Note, next, that agent a 's posterior odds conditional observing $C^* = \theta'$ are

$$\frac{a(\theta = \theta' | C^* = \theta')}{a(\theta = \theta'' | C^* = \theta')} = \frac{a(\theta = \theta')}{a(\theta = \theta'')} \frac{1 - \delta - \gamma + \gamma/|\Theta|}{\delta/(|\Theta| - 1) + \gamma/|\Theta|}.$$

Hence, for agents with

$$a(\theta')/a(\theta'') \leq \frac{1 - \delta - \gamma + \gamma/|\Theta|}{\delta/(|\Theta| - 1) + \gamma/|\Theta|} \quad (11)$$

for all $\theta', \theta'' \in \Theta$, the posterior mode conditional on observing $C^* = \theta$ always includes θ (where the mode may be set-valued). Hence, for these agents $R_a^*(c^*) = R_a(c^*) = \bar{R}(c^*)$, and for ρ sufficiently small that at least one agent of this type is included in $\Delta(\Theta) \setminus \mathcal{N}_\rho(a^*)$, we still have $\bar{R}^*(c^*) = \bar{R}(c^*)$. Note, moreover, that these are the only agents with $R_a^*(c^*) = \bar{R}(c^*)$, since if $a(\theta')/a(\theta'') > \frac{1 - \delta - \gamma + \gamma/|\Theta|}{\delta/(|\Theta| - 1) + \gamma/|\Theta|}$ for some $\theta', \theta'' \in \Theta$, then conditional on observing $C^* = \theta''$ agent a strictly prefers $d = \theta'$ to $d = \theta''$, which implies that $R_a^*(c^*) < R_a(c^*) = \bar{R}(c^*)$.

To complete the proof, let us consider the alternative rule c^{**} that always reports $C^{**} = X$. As noted in the proof of Claim 5, based on observing C^{**} agents can construct random variables with the same distribution as C^* , so $R_a^*(c^{**}) \leq R_a^*(c^*)$ for all a . Note, moreover, that for agents $a \in \Delta(\Theta) \setminus \mathcal{N}_\rho(a^*)$ whose priors satisfy (11), the posterior mode conditional on observing $C^{**} = \iota$ does not include all of Θ (since the uniform prior is excluded from $\Delta(\Theta) \setminus \mathcal{N}_\rho(a^*)$). Hence, these agents strictly prefer randomization over their (possibly set-valued) posterior mode to randomization over Θ conditional on $C^{**} = \iota$. Hence, $R_a^*(c^{**}) < R_a(c^*) = R_a^*(c^*)$ for these agents.

However, since these agents were the only ones with $R_a^*(c^*) = \overline{R}(c^*)$, this establishes that

$$\overline{R}(c^{**}) < \overline{R}^*(c^*) = \overline{R}(c^*),$$

as we wanted to show. \square

B Additional Results Referenced in the Text

B.1 Heterogeneous Loss Functions

In Section 2, we claimed that models with heterogeneous loss functions and homogeneous (or heterogeneous) priors can be cast into our setting. In this section we make this claim precise. Since the case with heterogeneous loss functions and heterogeneous priors nests the case with heterogeneous loss functions and homogeneous priors, we prove the result for the former case.

Specifically, suppose we begin with a model where agent a 's prior is again $\pi_a \in \Delta(\Theta)$, (where we no longer identify agents with their priors to allow the possibility that agents with the same prior have different loss functions) but agent a has loss function $L_a(d, \theta)$. Let us further suppose that $\sup_{a,d,\theta} L_a(d, \theta)$ is finite, and note that under this condition, since \mathcal{D} and Θ are both finite we can express $L_a(d, \theta)$ as

$$L_a(d, \theta) = \sum_{m=0}^M r_a(m) L_m(d, \theta)$$

where $L_0(d, \theta) = 0$, while for $m \geq 1$ the functions $L_m(d, \theta)$ are proportional to $1\{d = d_{j(m)}\} 1\{\theta = \theta_{k(m)}\}$, r_a satisfies $r_a(m) \geq 0$ for all m , and $\sum_{m=0}^M r_a(m) = 1$.

Using this observation, we can cast this example with heterogeneous priors into our baseline setting by augmenting the parameter space. Specifically, define $\Theta^* = \Theta \times \{0, \dots, M\}$, define agent a 's prior on this augmented space as $\pi_a^*(\theta^*) = \pi_a(\theta) r_a(m)$, and let all agents share the homogeneous loss function $L(d, \theta^*) = L_m(d, \theta)$.

Since the prior $\pi_a^*(\theta^*)$ imposes independence between θ and m , agent a 's posterior risk from action d conditional on (X, V) is

$$\mathbb{E}_a[L(d, \theta^*) | X, V] = \mathbb{E}_a \left[\sum_{m=1}^M r_a(m) L_m(d, \theta) | X, V \right] = \mathbb{E}_a[L_a(d, \theta) | X, V],$$

and so coincides with the posterior risk in the initial heterogeneous prior, heterogeneous loss model. However, since this holds for all (X, V) , the posterior risk conditional on (C, V) likewise coincides for all possible c . Hence, all risk calculations likewise coincide, and we have successfully recast the initial heterogeneous prior, heterogeneous loss model as a heterogeneous prior, homogeneous loss model. Whether the class of priors $\{\pi_a^* : a \in \mathcal{A}\}$ is closed and/or convex will depend on the original priors π_a and the functions r_a . Note, in particular, that in cases where the priors π_a are homogeneous in the original model, the class of constructed priors $\{\pi_a^* : a \in \mathcal{A}\}$ is closed and/or convex if and only if the same holds for $\{r_a : a \in \mathcal{A}\}$, and thus for the loss functions L_a .