A Model of Scientific Communication

Isaiah Andrews, *Harvard University and NBER*[*]

Jesse M. Shapiro, *Brown University and NBER*

February 2020

Abstract

We propose a model of empirical science in which an analyst makes a report to an audience after observing some data. Agents in the audience may differ in their beliefs or objectives, and may therefore update or act differently following a given report. We contrast the proposed model with a classical model of statistics in which the estimate directly determines the payoff. We identify settings in which the proposed model prescribes very different, and more realistic, optimal rules.

keywords: statistical communication, statistical decision theory

[*]This article previously circulated under the title "Statistical Reports for Remote Agents." We acknowledge funding from the National Science Foundation under Grant No. 1654234, the Alfred P. Sloan Foundation, the Silverman (1968) Family Career Development Chair at MIT, and the Brown University Population Studies and Training Center. 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, Xiaohong Chen, Max Kasy, Elliot Lipnowski, Adam McCloskey, Emily Oster, Mikkel Plagborg-Møller, Tomasz Strzalecki, and Neil Thakral. The paper also benefited from the comments of seminar and conference audiences at Brown University, Harvard University, the Massachusetts Institute of Technology, the University of Chicago, Boston College, Rice University, Texas A&M University, New York University, Columbia University, the University of California Los Angeles, the University of Texas at Austin, Oxford University, and from comments by conference discussant Jann Spiess. We thank our dedicated research assistants for their contributions to this project. E-mail: iandrews@fas.harvard.edu, jesse_shapiro_1@brown.edu.

1 Introduction

Statistical decision theory, following Wald (1950), is the dominant theory of optimal estimation in econometrics.[1] The classical decision-theoretic model envisions an analyst who estimates an unknown parameter based on some data. The performance of the estimate is judged by its proximity to the true value of the parameter. This judgment is formalized by treating the estimate as a decision that, along with the parameter, determines a realized payoff or loss. For example, if the loss is taken to be the square of the difference between the estimate and the parameter, then the expected loss of the decision-maker is the estimator's mean squared error, a standard measure of performance.

Although many scientific situations seem well described by the classical model, many others do not. Scientists often communicate their findings to a broad and diverse audience, consisting of many different agents (e.g., practitioners, policymakers, other scientists) with different opinions and objectives. These diverse agents may make different decisions, or form different judgments, following a given scientific report. In such cases, it is the beliefs and actions of these audience members which ultimately matter for realized payoffs or losses.

In this paper we propose an alternative model of empirical science that captures scientific situations of this kind. We develop sufficient conditions under which the proposed model predicts very different scientific reports from the classical model. We offer examples satisfying those conditions, and argue that in practical situations similar to these examples, scientists' reports seem to be more consistent with the predictions of the proposed model than with the predictions of the classical model.

In the proposed *communication model*, the analyst reports a recommended decision (or estimate) to an audience based on some observed data. Each agent in the audience takes her optimal decision (or forms an optimal estimate) after observing the analyst's report. Agents differ in their priors, and may therefore have different optimal decisions (or estimates) following a given report. To simplify the analysis, we assume that every possible prior belief on the unknown parameter is held by some agent in the audience. A reporting rule (specifying a distribution of recommendations

[1] See Lehmann and Casella (1998) for a textbook treatment of statistical decision theory and Stoye (2012) for a recent discussion of its relation to econometrics.

for each realization of the data) induces an expected loss for each agent, which we call the rule's *communication risk.*

We contrast the proposed communication model with a *decision model* in which the analyst selects a decision (or estimate) that directly determines the loss for all agents. Because agents in the decision model do not optimize following the analyst's report, each agent's expected loss, which we call the rule's *decision risk*, is weakly higher than under the communication model.[2] The implications of the decision model coincide with those of the classical model. Our analysis therefore proceeds by contrasting rankings of rules under the decision model with rankings under the communication model.

We find that the two models can imply very different rankings of rules. An example illustrates. Suppose that an analyst conducts a randomized controlled trial to assess the effect of a deworming medication on the average body weight of children in a low-income country. Even if deworming medication is known to (weakly) improve nutrition, sampling error means that the treatment-control difference may be negative. Under a canonical quadratic loss, the decision model implies that all audience members prefer that the analyst censor negative estimates at zero, since zero is closer to the (weakly positive) true effect than any negative number. Under the same loss, the communication model implies that censoring discards potentially useful information (the more negative the estimate, the weaker the evidence for a large positive effect), and has no corresponding benefit (agents can incorporate censoring when determining their optimal decisions or estimates). While a policymaker choosing a subsidy for the medication might well censor the decision (thus ruling out imposing a tax), we claim, and illustrate by example, that a scientist choosing a report for a research article would be unlikely to censor.

In the paper, we formalize and generalize the example. A rule is admissible with respect to a given definition of risk if no other rule yields a weakly lower risk for all agents and a strictly lower risk for at least one. While admissibility is a very weak notion of optimality, in settings like the deworming example there exists no rule that is simultaneously admissible for decision and communication. More generally, we show that the sets of decision-admissible and communication-admissible rules do

[2]Decision risk is what Lehmann and Casella (1998, Chapter 4) call the Bayes risk.

not intersect when (i) some decision is dominated in loss and (ii) the set of feasible decisions is smaller than the set of distinct optimal action profiles for agents.

We next turn to principles for choosing rules other than admissibility. One such principle is to minimize weighted average risk with respect to some weights on the audience. Applying this principle to decision risk leads to the class of Bayes decision rules. Bayes decision rules with respect to full-support priors have strong optimality properties in the classical setting, but because they are admissible in decision risk, they are inadmissible in communication risk under conditions (i) and (ii).

Another principle is to minimize the maximum risk over agents in the audience. In contrast to our findings for admissibility and optimality in weighted average risk, we find that any rule that is minimax in decision risk is minimax in communication risk. This finding establishes a sense in which any rule that is robust for decision-making is also robust for communication. However, minimax rules can be inadmissible, and under conditions (i) and (ii), any rule that is both minimax and admissible in decision risk is minimax and inadmissible in communication risk.

We apply our model to two types of scientific settings recently studied in econometrics: estimating the conditional expectation function when this function is known to be monotonically decreasing, and estimating the optimal choice of treatment from a finite set. In both settings, as in the deworming example, we show that the implications of the decision and communication models are different, and we argue that the communication model is more consistent with practice in some real-world situations.

Our analysis is positive in nature. We do not take a stand on which principles should guide scientists' choice of statistical reports. Nor do we argue that the communication model is a better description of all scientific situations than the decision model. Rather, we argue that the communication model better describes some important situations, and that in these situations, it seems to better match scientific practice.

Heterogeneity among agents plays a central role in our analysis. When agents are homogeneous, the distinction between decision and communication risk is inconsequential, because a benevolent analyst can simply report the agents' optimal decision (or estimate) given the data. When agents are instead heterogeneous, the distinction can be consequential, because different agents may prefer different decisions (or estimates). Although we model heterogeneity by assuming agents differ in their prior beliefs, we show that our model is equivalent to one in which agents instead differ in their loss functions.

Our baseline model assumes that the sets of possible parameter values, data realizations, and decisions are finite. This greatly simplifies the analysis but has the drawback that it precludes applying our framework to many canonical statistical situations such as estimating the mean of a Gaussian random variable. Online Appendix B.1 develops continuous versions of the example and applications discussed in the main text and shows that there remains a strong conflict between the decision and communication models, in the sense that changing the notion of risk can reverse dominance orderings.

We assume that the analyst's report takes the form of a recommended decision. This ensures that both decision and communication risk are well-defined for all rules we consider, and thus allows us to directly compare the ranking of rules under the two notions of risk. We discuss additional implications of this assumption in Section 2.3.

We are not aware of past work that studies the ranking of rules based on communication 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.[3] One conclusion from this literature is that when the analyst can communicate the full likelihood, data, or a sufficient statistic, the analyst should do so. Our analysis is particularly related to that of Hildreth (1963) who studies, among other topics, the properties of what we term communication risk in the single-agent setting. Banerjee et al. (forthcoming) study minimax experimental design when the audience has heterogeneous priors. Spiess (2018) studies optimal estimation in a setting where, unlike in our model, the objectives of the analyst and audience may be misaligned. Andrews et al. (2020) study the implications of communication risk for structural estimation

[3] See also Efron (1986) and Poirier (1988). A related literature (e.g., Pratt 1965, Kwan 1999, Abadie forthcoming, Abadie and Kasy 2019, Frankel and Kasy 2018) assesses the Bayesian interpretation of frequentist inference. Another literature (e.g., Zhang et al. 2013, Jordan et al. 2018, Zhu and Lafferty 2018a) considers the problem of distributing statistical estimation and inference across multiple machines when communication is costly.

in economics (see also Andrews et al. 2017).

Our setting is also related to the literature on comparisons of experiments following Blackwell (1951, 1953).[4] What we term communication 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 focus on the comparison between communication risk and decision risk for a given loss function and class of priors.

Our setting is broadly related to large literatures on strategic communication (Crawford and Sobel 1982) and information design (Bergemann and Morris 2019). As in Farrell and Gibbons (1989), the receivers (agents) in our setting are heterogeneous. As in Kamenica and Gentzkow (2011), the sender (analyst) in our 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 provides our main results on the implications of the communication model and its differences with the decision and classical models. Section 4 presents applications to two types of scientific situations. Section 5 concludes. An appendix following the body text provides proofs for the main results. Online Appendix A proves the results in the example and applications. Online Appendix B discusses extensions and auxiliary results.

2 Model

2.1 Primitives

An analyst 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 $\theta \in \Theta$. The analyst also observes a public random variable $V \sim U[0,1]$ that is independent of θ and X.

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

The analyst publicly commits to a rule $c : \mathcal{X} \times [0,1] \to \Delta(\mathcal{D})$ that maps from realizations of the data X and the public random variable V into a distribution over decisions $d \in \mathcal{D}$, for $\mathcal{D}$ a finite space. Let $\mathcal{C}$ denote the set of all such rules and with a slight abuse of notation let $c(X, V) \in \mathcal{D}$ denote the random realization from a given rule $c \in \mathcal{C}$.

Rules are evaluated by their performance with respect to a closed set $\mathcal{A} \subseteq \Delta(\Theta)$ of priors on the parameter space, which we will call the *audience*. Specifically, the function $\rho : \mathcal{C} \times \mathcal{A} \to \mathbb{R}_{\geq 0}$ describes the *risk* of rule $c \in \mathcal{C}$ with respect to any $a \in \mathcal{A}$. We discuss three possible choices of $(\mathcal{A}, \rho)$, along with their interpretation, in the next section.

The ordering of rules under the risk function $\rho(\cdot, \cdot)$ may depend on the prior $a \in \mathcal{A}$. We consider three canonical criteria by which the analyst could select a rule, or set of rules, in such a case.

Definition 1. *For a given risk function $\rho(\cdot, \cdot)$ and audience $\mathcal{A}$, a rule $c^* \in \mathcal{C}$ is*

- ***admissible*** *if there exists no rule $c \in \mathcal{C}$ such that $\rho(c, a) \leq \rho(c^*, a)$ for all $a \in \mathcal{A}$, with strict inequality for at least one $a \in \mathcal{A}$.*

- *ω-**optimal** if for a probability measure ω with support equal to $\mathcal{A}$*

$$\int_{\mathcal{A}} \rho(c^*, a)\, d\omega(a) = \inf_{c \in \mathcal{C}} \int_{\mathcal{A}} \rho(c, a)\, d\omega(a). \tag{1}$$

- ***minimax*** *if*

$$\sup_{a \in \mathcal{A}} \rho(c^*, a) = \inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} \rho(c, a).$$

These criteria are central to the classical study of statistics (Lehmann and Casella 1998). They also have an intuitive economic meaning. Admissibility excludes rules that are dominated. ω-optimality defines the set of rules optimal with respect to full-support Pareto weights ω on the audience. Minimaxity minimizes the worst-case risk and is familiar from welfare economics, game theory, and the study of decision-making under ambiguity. Existence of rules optimal according these criteria follows from standard arguments.

Proposition 1. *Suppose that $\rho(c, a)$ is continuous in a for all $c \in \mathcal{C}$, and that the set of risk functions $\{\rho(c, \cdot) : c \in \mathcal{C}\}$ is compact in the supremum norm. Then there exists an admissible rule, an ω-optimal rule for any ω, a minimax rule, and an admissible minimax rule. Moreover, any ω-optimal rule is admissible.*

Under the conditions of Proposition 1, the criteria we consider are available to the analyst, in the sense that they each admit at least one rule.

2.2 Risk Functions for Communication and Decision

In our setting, a model of empirical science is characterized by a risk function $\rho(\cdot, \cdot)$ and an audience $\mathcal{A}$, which together determine the implications of the canonical criteria for the analyst's choice of rules.

We focus on three such models.

Definition 2. *Fix a loss function $L : \mathcal{D} \times \Theta \to \mathbb{R}_{\geq 0}$. Then:*

- *The **communication model** takes $\mathcal{A} = \Delta(\Theta)$ and*

$$\rho(c, a) = R_a^*(c) = \mathrm{E}_a\left[\min_d \mathrm{E}_a\left[L(d, \theta) \,|\, c(X, V), V\right]\right],$$

  *for $R_a^*(c)$ the **communication risk** of rule $c \in \mathcal{C}$ for prior $a \in \Delta(\Theta)$ and E_a the expectation under a.*

- *The **decision model** takes $\mathcal{A} = \Delta(\Theta)$ and*

$$\rho(c, a) = R_a(c) = \mathrm{E}_a\left[L(c(X, V), \theta)\right]$$

  *for $R_a(c)$ the **decision risk** of rule $c \in \mathcal{C}$ for prior $a \in \Delta(\Theta)$.*

- *The **classical model** takes $\mathcal{A}$ to be the vertices of $\Delta(\Theta)$, such that each prior $a \in \mathcal{A}$ places probability 1 on some $\theta(a) \in \Theta$, and takes*

$$\rho(c, a) = R_{\theta(a)}(c) = \mathrm{E}_{\theta(a)}\left[L(c(X, V), \theta(a))\right]$$

  *for $R_{\theta(a)}(c)$ the **frequentist risk** of rule $c \in \mathcal{C}$ for the parameter $\theta(a) \in \Theta$.*

The communication risk $R_a^*(c)$ is the ex-ante expected loss under the prior a from taking the a-posteriori optimal decision after observing $c(X,V)$ and V. The decision risk $R_a(c)$ is the expected loss under the prior a from taking the decision $c(X,V)$. Finally, the frequentist risk $R_{\theta(a)}(c)$ is the expected loss under the parameter value $\theta(a)$ from taking the decision $c(X,V)$. Lemma 1 in the appendix shows that communication and decision risk satisfy the conditions of Proposition 1, while these conditions hold trivially for frequentist risk. Hence, Proposition 1 implies the existence of admissible, ω-optimal, and minimax rules in the communication, decision, and classical models.

The decision and classical models have different audiences but the same risk function. The following proposition shows that their implications coincide in a strong sense.

Proposition 2. *The decision model implies the same sets of admissible rules, ω-optimal rules for some ω, and minimax rules as the classical model.*

Intuitively, because the audience $\mathcal{A} = \Delta(\Theta)$ consists of the set of all possible convexifications of the vertices of $\Delta(\Theta)$, the dominance relation (admissibility), the implications of full-support averages (ω-optimality), and the worst-case risk (minimaxity) are all preserved when switching from the classical model's audience to the decision model's audience. This point is well-understood in the literature (see, e.g., Stoye 2011).

While we focus on the case with $\mathcal{A} = \Delta(\Theta)$ for simplicity of exposition, many of our results (though not Proposition 2) extend to the case of a general closed, convex audience $\mathcal{A} \subseteq \Delta(\Theta)$. We highlight these extensions in the proofs. That said, some degree of heterogeneity is necessary for the distinction between communication and decision risk to be interesting. For an audience $\mathcal{A} = \{a\}$ consisting of a single prior a, all three notions of optimality discussed in Definition 1 coincide, and optimality in the decision model implies optimality in the communication model.

It is worth briefly highlighting the role of randomization in the rules c. Specifically, the analyst may randomize publicly through the random variable V, considering $c : \mathcal{X} \times [0,1] \to \mathcal{D}$, privately through the choice of a randomized decision, $c : \mathcal{X} \to \Delta(\mathcal{D})$, or both publicly and privately, $c : \mathcal{X} \times [0,1] \to \Delta(\mathcal{D})$. For the decision and classical models private and public randomization are equivalent, in the sense that the risk depends only on the distribution of $c(X, V)|X$. For the communication model, by contrast, the nature of the randomization matters. Intuitively, since V is observed, publicly randomized rules $c : \mathcal{X} \times [0, 1] \to \mathcal{D}$ can be viewed as ex-ante mixtures over non-random rules $\tilde{c} : \mathcal{X} \to \mathcal{D}$, where it is revealed ex-post which rule $\tilde{c}$ was selected. By contrast, privately randomized rules $c : \mathcal{X} \to \Delta(\mathcal{D})$ can be viewed as ex-ante mixtures over non-random rules $\tilde{c} : \mathcal{X} \to \mathcal{D}$, where it is never revealed which $\tilde{c}$ was selected. Hence, the public random variable $V \sim U[0, 1]$ is a modeling device to allow the analyst to randomize across rules while revealing the outcome of the randomization to the audience. Lemma 2 in the appendix shows that public randomization always yields weakly lower communication risk than private randomization, with a strict inequality in some cases.

2.3 Interpretation of the Models

To interpret the models, we can identify each prior $a \in \Delta(\Theta)$ with some agent. Agents in the audience for a given scientific finding might include practitioners, policymakers, or other scientists. We highlight two different ways to interpret the decision $d \in \mathcal{D}$ and associated loss $L(d, \theta)$.

One interpretation is that the decision $d \in \mathcal{D}$ represents a real-world action whose consequences are captured by $L(d, \theta)$. For example, doctors may need to choose a treatment, policymakers to set a tax, and scientists to decide on what experiment to run next. On this interpretation, the decision model reflects a situation in which the analyst makes a decision on behalf of all agents, or equivalently, all agents are bound to take the decision recommended by the analyst. The communication model, by contrast, reflects a situation in which each agent is free to take her optimal decision given the information in the analyst's report.

Another interpretation is that the decision $d \in \mathcal{D}$ represents a best guess whose departure from the truth is captured by $L(d, \theta)$. This interpretation is evoked by canonical losses, such as $L(d, \theta) = (d - \theta)^2$, that increase in the distance between the estimate and the parameter. On this interpretation, the decision model reflects a situation in which each agent evaluates the quality of the analyst's guess according to the agent's prior. The communication model, by contrast, reflects a situation in which each agent evaluates the quality of the agent's own best guess, as informed by the analyst's report as well as the agent's prior.

In many real-world situations the agents in the audience for a given scientific finding will have diverse opinions and may therefore make different decisions, or form different best guesses about an unknown parameter, after learning the same report. The communication model better reflects such situations than does the decision model. In other situations—for example, a government committee deciding on the appropriate treatment to reimburse for a given diagnosis for all practitioners, or a scientific committee deciding where next to point a telescope that will provide data to many researchers—the decision model seems a better fit.

In our model, agents differ only in their prior beliefs about the parameter. Casual experience suggests that different agents (doctors, policymakers, scientists) do often disagree subjectively about how to interpret evidence, and the assumption of heterogeneous priors is one way to capture such disagreements (Morris 1995). But agents may also differ in their preferences or objectives. Online Appendix B.2 shows that models in which an audience of agents differ in their loss functions can be cast into our setting through appropriate relabeling and choice of $\mathcal{A}$.[5] The same appendix shows how to accommodate heterogeneous beliefs about the likelihood F_θ. Online Appendix B.3 deals with the case where an agent's payoff is determined by the agent's regret (i.e., loss or risk relative to a best-possible decision or rule).

It is also possible to do away with heterogeneous agents altogether by envisioning a single agent who receives some information about the parameter θ that is not available to the analyst at the time they make their report. Under this interpretation, $\mathcal{A}$ is the set of posterior beliefs that the agent may hold following receipt of the information. Under the decision model, the agent is bound by the recommended decision regardless of the additional information. Under the communication model, the agent is free to make an optimal decision after learning the additional information.

We assume throughout that the set of rules $\mathcal{C}$ available to the analyst is the set of estimators or decision rules, i.e. rules mapping from the sample space to distributions on the decision space $\mathcal{D}$. This assumption is important for comparing the implications

[5]Brown (1975) considers a setting with a collection of possible loss functions and proposes corresponding notions of admissibility.

of the decision and communication models, because decision risk is not well-defined for rules $c \notin \mathcal{C}$. Modifying the communication model to allow the use of a broader set of rules $\mathcal{C}' \supset \mathcal{C}$ would only strengthen the conflicts between decision and communication risk that we document in Section 3.[6] An unrealistic aspect of the communication model is that communication risk (unlike decision risk) is invariant to permutations of the analyst's report (e.g., reporting high values of d when the parameter θ is expected to be low, or vice versa). Online Appendix B.4 shows how to eliminate this unrealistic aspect of the model without altering our substantive findings.

In some situations, including some of the examples we discuss below, the communication constraint is not binding, in the sense that $\mathcal{D}$ is rich enough to allow the analyst to communicate all the decision-relevant information in the data. In general, this need not be the case, and when it is not, an analyst concerned with minimizing communication risk might like to use a richer vocabulary than $\mathcal{D}$. In practice, scientists often report estimates or other summaries that do not convey all of the information in the data. We think a plausible reason is that there are communication or information processing constraints on the part of the agents in the audience. Our model captures those constraints by the restriction to rules $c \in \mathcal{C}$. Restricting reports to use a finite vocabulary is a standard way to model such constraints in information theory (e.g., Cover and Thomas 2006), and has been explored in recent studies of nonparametric statistics under computational constraints (Zhu and Lafferty 2018a, b).

Online Appendix B.1 develops versions of our example and applications with continuous parameter spaces. For the example and one application we further allow for a continuous sample space $\mathcal{X}$ and decision space $\mathcal{D}$. With continuous $\mathcal{D}$, communication-optimal rules can take an unrealistic form, for example encoding the full data in the decimal expansion of $c(X, V)$. The appendix therefore focuses on showing that the decision and communication models deliver conflicting rankings of particular, realistic-seeming rules. How best to model communication constraints for general settings with continuous decision spaces seems a potentially interesting topic for future work.

[6] If, according to a given optimality criterion, any rule that is optimal in decision risk is not optimal in communication risk, then the same applies if we evaluate communication risk with respect to $\mathcal{C}' \supset \mathcal{C}$.

3 Implications of the Communication Model

In this section we discuss the implications of the communication model, emphasizing the contrast with the decision (and hence classical) model. To illustrate the key intuitions, we return to (and elaborate) the example from the introduction.

Example. An analyst observes data on weight gain for a sample of n children enrolled in a randomized trial of deworming drugs. Weight is measured to finite precision, so the weight gain for child i is $X_i \in \mathcal{X}_0$ for $\mathcal{X}_0$ a finite set with $|\mathcal{X}_0| \geq 2$. The data are thus

$$X = (X_1, ..., X_n) \in \mathcal{X} = \mathcal{X}_0^n.$$

Indices i are assigned at random, with children $\{1, ..., \underline{n}\}$ assigned to control and children $\{\underline{n}+1, ..., n\}$ assigned to treatment, with $\underline{n}$, $n - \underline{n} \geq 3$. Children's weight gains X_i, X_j are independent for all $i \neq j$. For children in the control group, $X_i \sim F_0(\underline{\theta})$ and for those in the treatment group $X_i \sim F_0\left(\overline{\theta}\right)$, where $F_0(t)$ is an exponential family with probability mass function of the form $f_0(x; t) = \exp(tx)\, h(t)\, g(x)$. This implies that the control and treatment group means

$$\left(\underline{X}, \overline{X}\right) = \left(\frac{1}{\underline{n}} \sum_{i=1}^{\underline{n}} X_i, \frac{1}{n - \underline{n}} \sum_{i=\underline{n}+1}^{n} X_i\right)$$

are a sufficient statistic for $\theta = \left(\underline{\theta}, \overline{\theta}\right)$. Let $\underline{\mathcal{X}}_M$ and $\overline{\mathcal{X}}_M$ denote the sets of possible values for each of these means.

The average treatment effect of deworming drugs on child weight is

$$\mathrm{E}_{F_0\left(\overline{\theta}\right)}[X_i] - \mathrm{E}_{F_0(\underline{\theta})}[X_i] = ATE(\theta),$$

which can be estimated without bias by the difference in means $\overline{X} - \underline{X}$. Suppose that the average treatment effect is known *a priori* to be non-negative, and that $\underline{\theta}, \overline{\theta} \in \Theta_0$ for Θ_0 a finite set with $|\Theta_0| \geq 2$ and $\max_{\overline{\theta} \in \Theta_0} \mathrm{E}_{F_0\left(\overline{\theta}\right)}[X_i] - \min_{\underline{\theta} \in \Theta_0} \mathrm{E}_{F_0(\underline{\theta})}[X_i] > 2/3$. The parameter space is thus

$$\Theta = \left\{\theta \in \Theta_0^2 : ATE(\theta) \geq 0\right\} = \left\{\theta \in \Theta_0^2 : \overline{\theta} \geq \underline{\theta}\right\}.$$

The audience consists of governments who must decide how much to subsidize (or tax) deworming drugs. The optimal Pigouvian subsidy is simply $ATE(\theta)$. The governments face a loss $L(d,\theta) = (d - ATE(\theta))^2$ for d the per-unit subsidy, with $d < 0$ denoting a tax. The set of feasible decisions,

$$\mathcal{D} = \left\{\overline{x} - \underline{x} : (\underline{x}, \overline{x}) \in \underline{\mathcal{X}}_M \times \overline{\mathcal{X}}_M\right\},$$

corresponds to the support of the average treatment effect estimate $\overline{X} - \underline{X}$. Online Appendix B.5 provides a microfoundation for the quadratic loss $(d - ATE(\theta))^2$ in a Pigouvian setting.

Because $\overline{\theta} \geq \underline{\theta}$ by assumption, a tax ($d < 0$) is never optimal. Therefore, rules that sometimes recommend $d < 0$ are unappealing from the standpoint of decision risk. Nevertheless, because the sample is finite we may have $\overline{X} < \underline{X}$ even though $\overline{\theta} \geq \underline{\theta}$. Agents who are, say, unsure whether to subsidize a little or not at all may value knowing that $\overline{X} < \underline{X}$, because they may apply a lower subsidy when $\overline{X} < \underline{X}$ than when, say, $\overline{X} = \underline{X}$. Therefore, from the standpoint of communication risk, rules that never report $d < 0$ may be unappealing because they suppress useful information.

We may alternatively envision the loss as capturing the scientific community's desire for a good guess of the true average treatment effect. On this interpretation, a guess $d < 0$ is again unappealing from the standpoint of decision risk (such a guess cannot be right), but may be appealing from the standpoint of communication risk (because it conveys useful information that agents can use in formulating their own guesses). ▲

The Example illustrates a setting in which we expect the implications of the communication and decision models to be very different. We next formalize the source of the tension. The following subsections show how it manifests under each of the criteria that we consider for choosing rules.

Definition 3. *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 $\theta' \in \Theta$.*

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 some parameter value.

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,*

$$\mathcal{P}^* = \left\{ P \in \mathcal{P} : \left\{ \cap_{X \in \mathcal{X}_p} \arg\min_{d \in \mathcal{D}} \mathrm{E}_a \left[L\left(d, \theta\right) | X \right] \right\} \neq \emptyset \text{ for all } \mathcal{X}_p \in P,\, a \in \Delta\left(\Theta\right) \right\}.$$

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

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

The effective size of the sample space is the smallest number of cells into which we can partition the sample space $\mathcal{X}$ such that knowing only which cell contains X is sufficient for all agents to take optimal decisions. For example, such a partition would group data realizations x and x' that imply the same likelihood, i.e., for which $\Pr_\theta\left\{X = x\right\} = \Pr_\theta\left\{X = x'\right\}$ for all θ.

Claim 1. In the Example, any $d < 0$ is dominated in loss, and $N\left(\mathcal{X}\right) = \left|\underline{\mathcal{X}}_M\right| \times \left|\overline{\mathcal{X}}_M\right| > |\mathcal{D}|$.

Intuitively, any decision $d < 0$ is dominated by the decision $d' = 0$ because decision d' is strictly closer to the optimal decision $ATE\left(\theta\right)$. At the same time any two distinct realizations of the sufficient statistics imply distinct optimal actions for some agent, so the effective size of the sample space is $\left|\underline{\mathcal{X}}_M\right| \times \left|\overline{\mathcal{X}}_M\right|$.

3.1 Admissibility

When there exists a dominated decision and $\mathcal{N}\left(\mathcal{X}\right) \geq |\mathcal{D}|$, the communication and decision models have non-overlapping sets of admissible rules.

Proposition 3. *If (i) there exists a decision $d \in \mathcal{D}$ that is dominated in loss and (ii) $\mathcal{N}\left(\mathcal{X}\right) \geq |\mathcal{D}|$, then any rule $c \in \mathcal{C}$ that is admissible under the decision model is inadmissible under the communication model, and vice versa.*

The proof of Proposition 3 proceeds in two steps. The first step establishes that, under the decision model, if $d \in \mathcal{D}$ is dominated in loss by $d' \in \mathcal{D}$, then any rule

$c \in \mathcal{C}$ that recommends d with positive probability is dominated by the rule $c' \in \mathcal{C}$ that instead recommends d'.

The second step establishes that, under the communication model, if the effective size of the sample space is weakly larger than the decision space, then any rule $c \in \mathcal{C}$ which never uses some $d \in \mathcal{D}$ is dominated by another rule $c' \in \mathcal{C}$ which sometimes uses d. The reason is that, for any such c, 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 $X \in \mathcal{X}^*$. It follows that c can be dominated by a rule c' that reports d on the subset of $\mathcal{X}^*$ where a would like to take a particular action, which strictly reduces the communication risk for agent a without increasing it for any other agent.

Propositions 2 and 3 immediately establish conditions for a conflict between the communication and classical models:

Corollary 1. *If (i) there exists a decision $d \in \mathcal{D}$ that is dominated in loss and (ii) $\mathcal{N}(\mathcal{X}) \geq |\mathcal{D}|$, then any rule $c \in \mathcal{C}$ that is admissible under the classical model is inadmissible under the communication model, and vice versa.*

Claim 1 allows us to apply Proposition 3 and Corollary 1 to the Example.

Corollary 2. *In the Example, there is no rule $c \in \mathcal{C}$ that is admissible under both the decision model and the communication model, and no rule $c \in \mathcal{C}$ that is admissible under both the classical model and the communication model.*

Online Appendix B.1.1 develops a variant of this example with Gaussian data and a continuous parameter space, and shows that decision and communication risk induce conflicting dominance orderings.

It is also possible to modify the Example so the analyst reports estimates of the treatment and control group means rather than of their difference. With this modification the communication constraint is not binding, in the sense that it is feasible to report a sufficient statistic for θ. The hypotheses of Proposition 3, and hence the conclusions of Corollary 2, continue to hold for this alternative formulation.[7]

[7]Formally, let $d = (\underline{d}, \overline{d}) \in \mathcal{D} = \underline{\mathcal{X}}_M \times \overline{\mathcal{X}}_M$, and define the loss as

$$L(d, \theta) = \left(\overline{d} - \underline{d} - ATE(\theta)\right)^2.$$

Example. *(Continued)* Kruger et al. (1996) conducted an early randomized controlled trial of the effect of anthelmintic therapy (deworming drugs) on children's growth. A separate randomization was used to study the effect of iron-fortified soup. Children were weighed to the nearest 0.1kg at baseline and endline. Among children who received unfortified soup, those receiving deworming drugs had a *lower* average growth over the intervention period (mean weight gain of 0.9kg, $n = 15$) than those receiving a placebo treatment (mean weight gain of 1.0kg, $n = 14$; see Table 4 of Kruger et al. 1996). Kruger et al. (1996) state that "[Positive effects on weight gain] can be expected with reduction in diarrhoea, anorexia, malabsorption, and iron loss caused by parasitic infection" (p. 10). In a later review of the literature, Croke et al. (2016) state that "there is no scientific reason to believe that deworming has negative side effects on weight" (p. 19). If we interpret these statements to mean that the average treatment effect is known to be non-negative, then censoring the estimated treatment effect at 0 (i.e. reporting that the treatment and control groups experienced the same average weight gain) would lead to an estimate strictly closer to the truth than the negative estimate implied by the group means, and would therefore dominate in mean squared error. However, Kruger et al. (1996) report the group means and do not report a censored estimate. Indeed, among the four studies that Croke et al. (2016) identify as implying negative point estimates of the effect of deworming drugs on weight, none published a censored point estimate.[8]

The preceding analysis suggests one possible explanation. While each member of the audience for the research might be happy with *some* censoring scheme, different members of the audience would like different ones, and (under the conditions above)

By Claim 1, $\overline{d} - \underline{d} < 0$ is dominated in loss and $N(\mathcal{X}) = |\underline{\mathcal{X}}_M| \times |\overline{\mathcal{X}}_M| \geq |\mathcal{D}|$.

[8]Croke et al. (2016, Figure 2) identify 4 negative point estimates out of a total of 22 reviewed. These 4 negative point estimates are from 4 distinct studies (including Kruger et al. 1996), out of a total of 20 distinct studies reviewed. Donnen et al. (1998, Table 2) report the regression-adjusted weight gains for a group treated with mebendazole and a control. They further report that the treated group's gain is statistically significantly below that of the control group at all time horizons considered. Croke et al. (2016, Figure 2) report a statistically significant effect on weight gain of -0.45kg based on the data from Donnen et al. (1998). Miguel and Kremer (2004, Table V) report treatment and control group means of standardized weight-for-age and a statistically insignificant difference in means of -0.00 to rounding precision. Croke et al. (2016, Figure 2) report a statistically insignificant effect on weight of -0.76kg based on the data from Miguel and Kremer (2004). Awasthi et al. (2000, Table 1) report treatment and control group means of weight gain and report that these are not statistically different. Croke et al. (2016, Figure 2) report a statistically insignificant effect of -0.05kg based on the data from Awasthi et al. (2000).

any single censoring scheme sacrifices decision-relevant information for some audience member. ▲

3.2 ω-optimality

In the classical setting, a rule c is a Bayes decision rule (or Bayes estimator) if it prescribes the optimal decision given the data X for a Bayesian agent with some proper prior (e.g., Lehmann and Casella 1998, p. 6; Robert 2007, p. 63). Such decision rules have strong optimality properties in the classical setting. In particular, Complete Class Theorems show that in many settings any rule that cannot be expressed as Bayes is dominated by one that can be. These Theorems have been invoked as part of the justification for using Bayes procedures (e.g., Robert 2007, p. 512).

Under the decision model any rule that is ω-optimal is a Bayes decision rule in the classical sense. The reason is that the weighted average decision risk for the audience can be expressed as the decision risk for a single agent with a weighted average prior $a_\omega(\theta) = \int_{\Delta(\Theta)} a(\theta)\, d\omega(a)$:

$$\int_{\Delta(\Theta)} R_a(c)\, d\omega(a) = \sum_{\theta\in\Theta} \left(\mathrm{E}_\theta \left[L\left(c\left(X,V\right),\theta\right)\right] \int_{\Delta(\Theta)} a(\theta)\, d\omega(a) \right). \tag{2}$$

Unlike decision risk, communication risk does not generally admit a representation like (2). Nonetheless, we show in Online Appendix B.6 that a Complete Class Theorem holds in both the decision model and the communication model.[9]

Propositions 1 and 3 establish conditions under which the set of ω-optimal rules in the communication model does not overlap with the set of ω-optimal rules in the decision model or with the set of Bayes decision rules with respect to full-support priors.

Corollary 3. *If (i) there exists a decision $d \in \mathcal{D}$ that is dominated in loss and (ii) $\mathcal{N}(\mathcal{X}) \geq |\mathcal{D}|$, then:*

[9]To describe this result, define $\underline{\omega}$-optimality analogously to ω-optimality, except that the support of $\underline{\omega}$ may be a strict subset of $\Delta(\Theta)$. Unlike ω-optimality, $\underline{\omega}$-optimality need not imply admissibility. The Complete Class Theorem states that any rule that is not $\underline{\omega}$-optimal is dominated by some $\underline{\omega}$-optimal rule.

(a) Any rule $c \in \mathcal{C}$ that is ω-optimal under the decision model is inadmissible, and therefore not ω'-optimal for any ω', under the communication model, and vice versa.

(b) Any rule $c \in \mathcal{C}$ that can be expressed as a Bayes decision rule with respect to a prior with full support on Θ is inadmissible, and therefore not ω-optimal for any ω, under the communication model.

(c) Any rule $c \in \mathcal{C}$ that is ω-optimal under the communication model is not an optimal decision rule for any agent with a full-support prior, $a \in \operatorname{int}(\Delta(\Theta))$.

Corollary 3 implies a very strong sense in which the communication model predicts different ω-optimal rules than the decision (and hence classical) model. This arises because (under the conditions of the corollary) Bayes decision rules with respect to full-support priors discard actionable information.

Example. *(Continued)* Suppose that Kruger et al. (1996) had selected a full-support prior $a \in \operatorname{int}(\Delta(\Theta))$ and reported the associated Bayes decision. By construction, under both the decision and communication models, this rule is optimal for audience members with prior a. Moreover, under the decision model, the rule is optimal with respect to some full-support Pareto weights ω on the audience, and hence admissible. Under the communication model, however, this rule is not optimal with respect to any such Pareto weights, and is inadmissible. The reason is that any such Bayes decision rule never selects $d < 0$, and so censors the data unnecessarily.

Figure 1 illustrates this intuition in a toy numerical example with $F_0(t)$ a Bernoulli distribution with success probability t, $\Theta_0 = \{0.3, 0.7\}$, and hence $\Theta = \{(0.3, 0.3), (0.3, 0.7), (0.7, 0.7)\}$. The upper-left plot shows the Bayes decision rule $\tilde{c}$ for a particular full-support prior $\tilde{a} \in \operatorname{int}(\Delta(\Theta))$. The rule $\tilde{c}$ never makes a negative report, and is therefore dominated in communication risk by another rule $\tilde{c}'$ that does sometimes make a negative report, shown in the upper-right plot. The lower plot shows, for each a, the normalized difference $R_a^*(\tilde{c}) - R_a^*(\tilde{c}')$ in communication risk between the rule $\tilde{c}$ and the dominating rule $\tilde{c}'$.

The dominating rule $\tilde{c}'$ achieves weakly lower communication risk than the rule $\tilde{c}$ for all agents $a \in \Delta(\Theta)$. Because rule $\tilde{c}$ is the Bayes decision rule for agent $\tilde{a}$, the dominating rule $\tilde{c}'$ achieves the same communication risk as $\tilde{c}$ for this agent. The same holds for other agents who, like agent $\tilde{a}$, put a lot of prior mass on the parameter values $\{(0.3, 0.3), (0.7, 0.7)\}$ under which the intervention is ineffective. However, for other agents, such as $\tilde{a}'$, who put more prior mass on the possibility that the intervention is effective, the dominating rule $\tilde{c}'$ achieves strictly lower communication risk than the rule $\tilde{c}$. Intuitively, such agents value knowing when $\overline{X} < \underline{X}$, i.e. when the evidence for gains from treatment is especially weak.

Croke et al. (2016, Figure 2) review 20 distinct studies reporting on randomized controlled trials of the effects of deworming drugs on children's growth. None of these reports a Bayesian posterior mean for a proper prior, or other explicitly Bayesian estimate of the effect of deworming on weight. Our analysis illustrates one possible reason for this, which is that reporting a Bayes decision with respect to one prior can sacrifice information useful to an agent with a different prior. ▲

3.3 Maximum Risk

The set of minimax rules in the communication model nests that in the decision model.

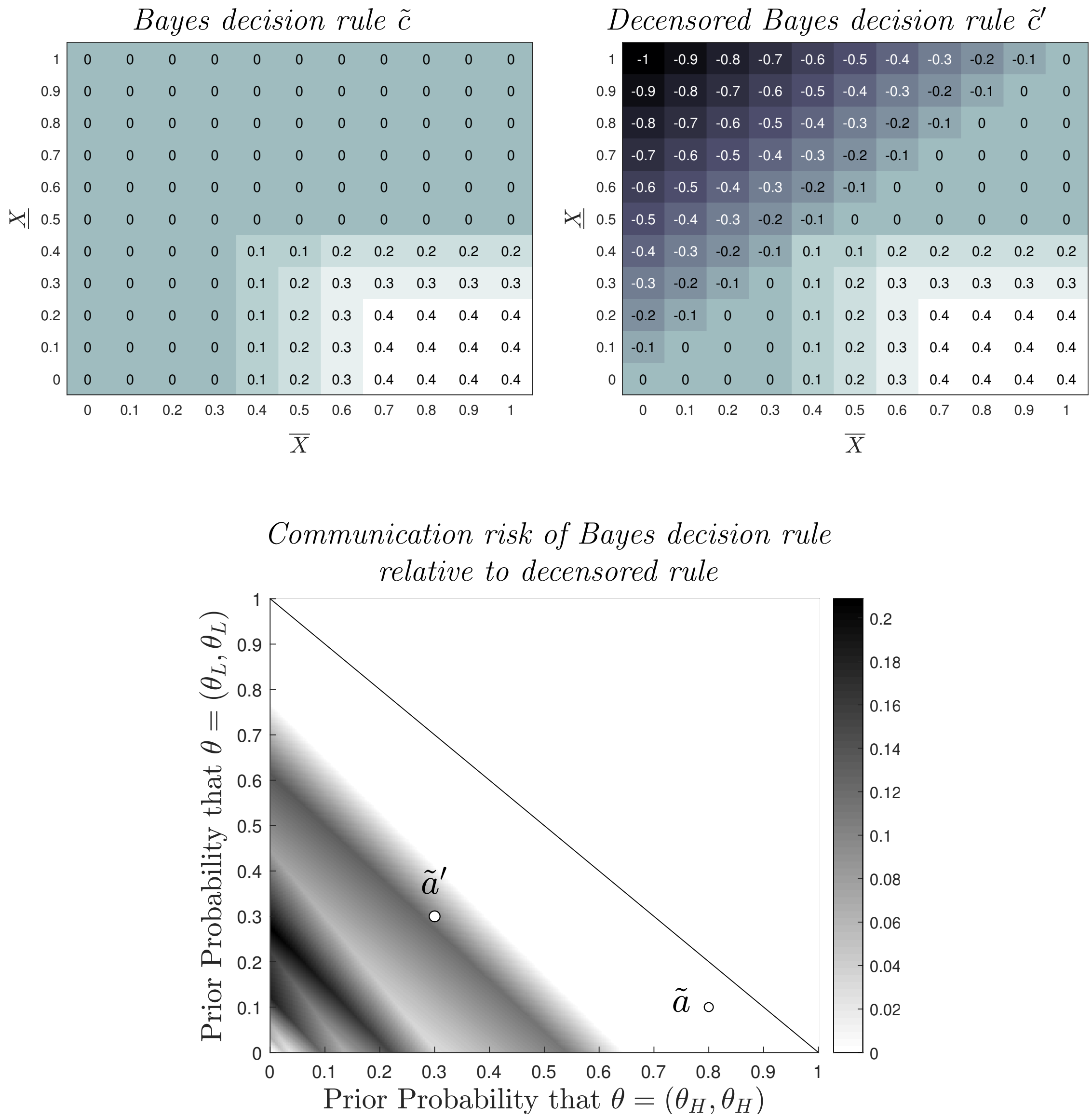
Theorem 1. *Any rule $c^* \in \mathcal{C}$ that is minimax under the decision (or classical) model is minimax under the communication model.*

An intuition for the proof is as follows. Pick some rule $c^* \in \mathcal{C}$ that is minimax in decision risk. Because the set of priors $\Delta(\Theta)$ is convex, we can show (following results in Grünwald and Dawid 2004, who in turn build on the classic minimax theorem of von Neumann 1928) that there exists some worst-off agent $a^* \in \Delta(\Theta)$ for whom c^* is optimal in decision risk. For this agent a^*, it follows that c^* must also be optimal in communication risk, because the agent's optimal decision following any report $c^*(X, V)$ will simply be the report itself. But because this agent a^* is worst-off in decision risk, it follows that this same agent must also be worst-off in communication risk. Collecting this reasoning gives:

$$\inf_{c \in \mathcal{C}} \sup_{a \in \Delta(\Theta)} R_a(c) = \inf_{c \in \mathcal{C}} R_{a^*}(c) = R_{a^*}(c^*) = R^*_{a^*}(c^*) = \inf_{c \in \mathcal{C}} R^*_{a^*}(c) = \inf_{c \in \mathcal{C}} \sup_{a \in \Delta(\Theta)} R^*_a(c).$$

Hence c^* is minimax in communication risk. Convexity of the audience is important

Figure 1: Illustration of a Bayes decision rule in running example



Note: The plots illustrate properties of a Bayes decision rule in the running example with $\Theta_0 = \{0.3, 0.7\}$ and thus $\Theta = \{(0.3, 0.3), (0.3, 0.7), (0.7, 0.7)\}$. We set $n = 20$, $\underline{n} = 10$, and $\mathcal{X}_0 = \{0, 1\}$, so that $\underline{\mathcal{X}}_M = \overline{\mathcal{X}}_M = \{0, 0.1, ..., 1\}$ and $\mathcal{D} = \{-1, -0.9, ..., 1\}$. We assume that $F_0(t)$ is a Bernoulli distribution with success probability t. The upper left plot shows the Bayes decision rule $\tilde{c}$ for a particular full-support prior $\tilde{a} \in \text{int}(\Delta(\Theta))$, specifically $\tilde{a} = (0.1, 0.1, 0.8)$. The upper right plot shows a rule $\tilde{c}'$ such that $\tilde{c}'(X) = \tilde{c}(X) + 1\{\tilde{c}(X) = 0\} \min\{0, \overline{X} - \underline{X}\}$. The lower plot shows the difference in communication risk $R_a^*(\tilde{c}) - R_a^*(\tilde{c}')$ between the two rules for each agent $a \in \Delta(\Theta)$, normalized by the maximum communication risk (over all agents) from a report of the full data. In the plot, we label the prior $\tilde{a}$, for which $R_{\tilde{a}}^*(\tilde{c}) = R_{\tilde{a}}^*(\tilde{c}')$, as well as another prior $\tilde{a}'$, for which $R_{\tilde{a}'}^*(\tilde{c}) > R_{\tilde{a}'}^*(\tilde{c}')$.

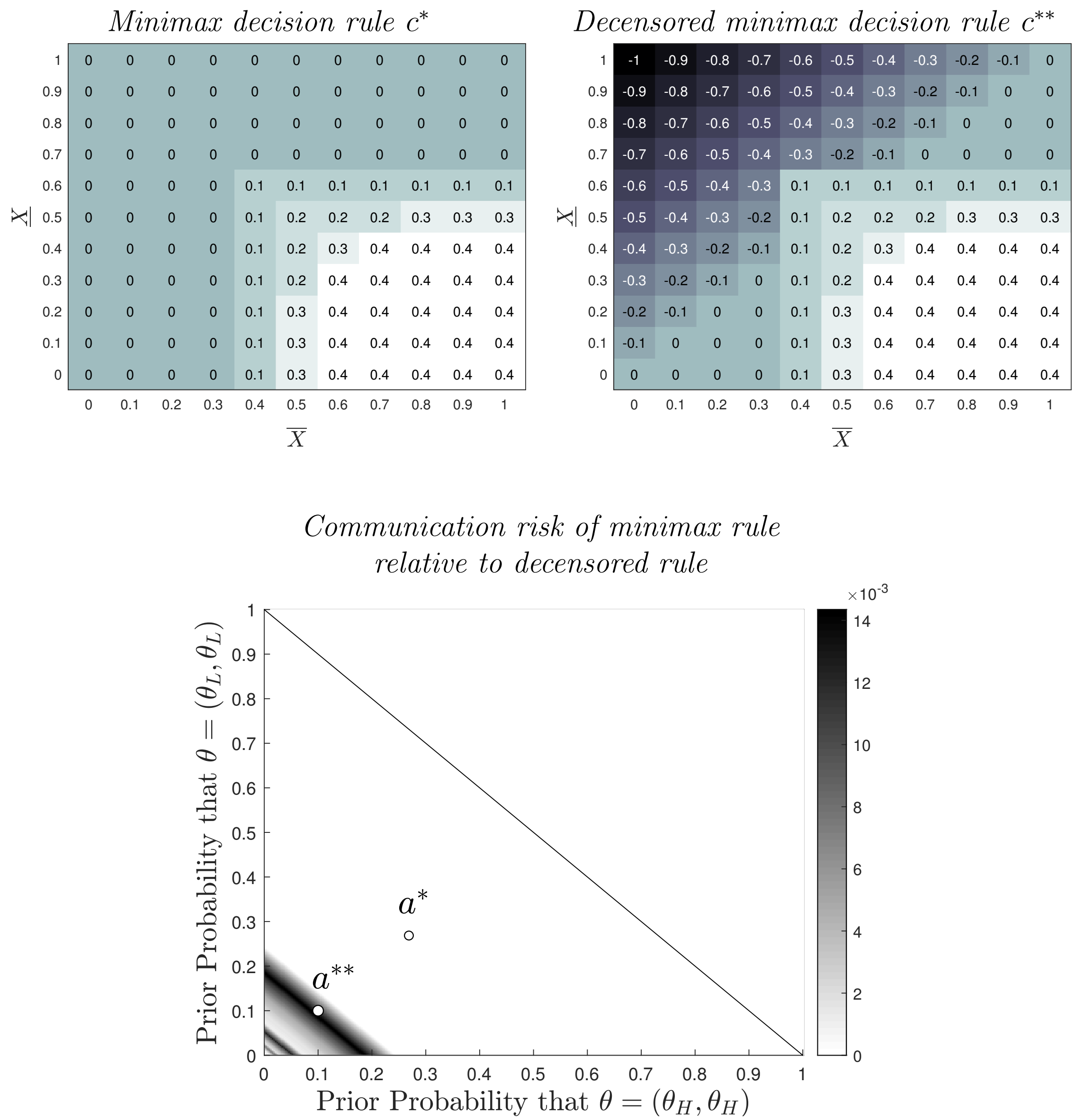
for this line of reasoning, because without it we cannot guarantee that the worst-case prior corresponds to an agent a^* in the audience.

Although minimax rules under the decision model are guaranteed to be minimax under the communication model, they need not be appealing if the analyst cares about more than the worst-case risk. Recall that by Proposition 1 and Lemma 1, both the decision and communication models admit a nonempty set of minimax admissible rules. Under the conditions of Proposition 3, these sets do not intersect, and thus there is no rule that is minimax admissible under both models.

Corollary 4. *If (i) there exists a decision $d \in \mathcal{D}$ that is dominated in loss and (ii) $\mathcal{N}(\mathcal{X}) \geq |\mathcal{D}|$, then any rule $c^* \in \mathcal{C}$ that is minimax and admissible under the decision (or classical) model is minimax and inadmissible under the communication model.*

Example. *(Continued).* In the Example, any rule c^* that is minimax under the decision model reports $d \geq 0$ with probability one. Hence, all rules that are minimax under the decision model are inadmissible under the communication model in this setting. Figure 2 continues our numerical illustration. The upper-left plot shows a minimax decision rule c^* in this numerical example. The rule c^* never makes a negative report, and is therefore dominated in communication risk by another rule c^{**} that does sometimes make a negative report, shown in the upper-right plot. The lower plot shows, for each a, the normalized difference $R_a^*(c^*) - R_a^*(c^{**})$ in communication risk between the rule c^* and the dominating rule c^{**}. The worst-off agent a^*, and other agents who likewise put significant mass on parameters under which the treatment is ineffective, do not benefit in communication risk from the dominating rule c^{**}. However, agents who, like the labeled agent a^{**}, believe a priori that the treatment is very likely effective, benefit in communication risk from rule c^{**} because overturning their priors requires strong evidence of ineffectiveness. Such evidence can be provided by reporting the treatment-control difference $\overline{X} - \underline{X}$ in cases where this difference is negative. ▲

Figure 2: Illustration of minimax decision rule in running example



Note: The plots illustrate properties of a minimax decision rule in the running example with $\Theta_0 = \{0.3, 0.7\}$ and thus $\Theta = \{(0.3, 0.3), (0.3, 0.7), (0.7, 0.7)\}$. We set $n = 20$, $\underline{n} = 10$, and $\mathcal{X}_0 = \{0, 1\}$, so that $\underline{\mathcal{X}}_M = \overline{\mathcal{X}}_M = \{0, 0.1, ..., 1\}$ and $\mathcal{D} = \{-1, -0.9, ..., 1\}$. We assume that $F_0(t)$ is a Bernoulli distribution with success probability t. The upper left plot shows the minimax decision rule c^*. The upper right plot shows a rule c^{**} such that $c^{**}(X) = c^*(X) + 1\{c^*(X) = 0\} \min\{0, \overline{X} - \underline{X}\}$. The lower plot shows the difference in communication risk $R_a^*(c^*) - R_a^*(c^{**})$ between the two rules for each agent $a \in \Delta(\Theta)$, normalized by the maximum communication risk (over all agents) from a report of the full data. In the plot, we label the prior a^* of the worst-off agent, for which $R_{a^*}^*(c^*) = R_{a^*}^*(c^{**})$, as well as another prior a^{**}, for which $R_{a^{**}}^*(c^*) > R_{a^{**}}^*(c^{**})$.

4 Applications

In this section we apply our model to two types of scientific situations studied in the recent econometrics literature. In both situations we show that the implications of the decision and communication models are different, and we argue that the communication model is more consistent with some observed scientific practices.

4.1 Estimation of a Monotone Function

Our first application is to the estimation of a function that is known to be monotonically decreasing, which is a special case of estimation under shape restrictions. Shape restrictions arise in many important economic situations, and econometric procedures that exploit these restrictions can offer improved performance according to conventional criteria (see, e.g., the review by Chetverikov et al. 2018).

As a concrete example, consider estimating the effect of the length of a job-seeker's current spell of unemployment on the probability that her resume receives a callback. Following Oberholzer-Gee (2008), Kroft et al. (2013), and Eriksson and Rooth (2014), we may imagine that the analyst conducts an audit study, submitting artificial resumes with randomly assigned employment histories, and recording the responses from potential employers. Multiple classes of economic models (reviewed, for example, in Kroft et al. 2013, Section II) imply that a longer current unemployment spell makes an applicant less attractive to an employer, which can be interpreted as a monotonicity restriction that the conditional probability of a callback is weakly decreasing in the duration of the current unemployment spell.

Formally, let $Y_i \in \{0,1\}$ denote a binary outcome, and suppose we are interested in the conditional mean $\mathrm{E}[Y_i|Z_i]$ of Y_i given some discrete $Z_i \in \mathcal{Z} = \{z_1, ..., z_J\}$, with $z_1 < z_2 < ... < z_J$. In the resume audit setting, we can think of Y_i as an indicator for whether resume i generates a callback, and Z_i as the duration (in months) of the fictitious applicant's current unemployment spell.

The analyst observes $n \geq 3$ independent draws of Y_i for each predictor value z_j. Denote the fraction of successes when $Z_i = z_j$ by $X_j \in \left\{0, \frac{1}{n}, ..., 1\right\}$. The number of successes nX_j follows a binomial distribution and it is without loss to represent the data as the vector of success rates,

$$X = (X_1, ..., X_J) \in \mathcal{X} = \left\{0, \frac{1}{n}, ..., 1\right\}^J.$$

The unknown parameter is the vector $\theta = (\theta_1, ..., \theta_J)$ of conditional means $\theta_j = \mathrm{E}[Y|Z = z_j]$. We assume that conditional mean is weakly decreasing in j

$$\theta_1 \geq \theta_2 \geq ... \geq \theta_J. \tag{3}$$

We assume that $\theta_j \in \Theta_0 \subset (0, 1)$ for each j, for Θ_0 a finite set such that $\max\{\Theta_0\} - \min\{\Theta_0\} > 2/3$. The parameter space is thus

$$\Theta = \left\{\theta \in \times_{j=1}^{J} \Theta_0 : \theta_1 \geq \theta_2 \geq ... \geq \theta_J\right\}.$$

We take the decision space to equal the sample space, $\mathcal{D} = \mathcal{X}$, so the communication constraint is not binding in the sense that it is feasible to report the full data X. Finally, we suppose that the objective is to estimate θ, which we formalize by considering the quadratic loss

$$L(d, \theta) = \|d - \theta\|_2^2 = \sum_j (d_j - \theta_j)^2.$$

The results of this section hold for any loss of the form $L(d, \theta) = \|d-\theta\|_p^p$ or $L(d, \theta) = \|d - \theta\|_p$ for $p > 1$.

A natural rule is to simply report the data, $c(X, V) = X$. All audience members agree that such a report is unbiased, in the sense that $\mathrm{E}_a[X - \theta] = 0$ for all $a \in \Delta(\Theta)$. However, X_j need not be decreasing in j, and indeed $\Pr_\theta\{X_j > X_{j-1}\} > 0$ for all j and all $\theta \in \Theta$. Hence, this rule does not respect the restrictions on the parameter space. Let $d_{(j)}$ sort the elements of d in decreasing order, so $d_{(1)}$ is the largest and $d_{(J)}$ the smallest. Define

$$d_j^*(d) = d_{(j)}$$

as the decision which sorts the elements in d in decreasing order. The following result is immediate from Chernozhukov et al. (2009).

Claim 2. Any decision d with $d_j > d_{j-1}$ for some j is dominated in loss by $d^*(d)$.

At the same time, one can show that any two distinct realizations of the data imply distinct optimal actions for some agent, which implies the following.

Claim 3. The effective size of the sample space is equal to the size of the sample and decision spaces, $N(\mathcal{X}) = |\mathcal{X}| = |\mathcal{D}|$.

It follows by Proposition 3 that there is no rule $c \in \mathcal{C}$ that is admissible under both the decision model and the communication model, and no rule $c \in \mathcal{C}$ that is admissible under both the classical model and the communication model. Online Appendix B.1.2 develops a version of this example with X normally distributed and a continuous parameter space $\Theta_0 \subseteq \mathbb{R}$, and shows that decision and communication risk again imply different dominance orderings.

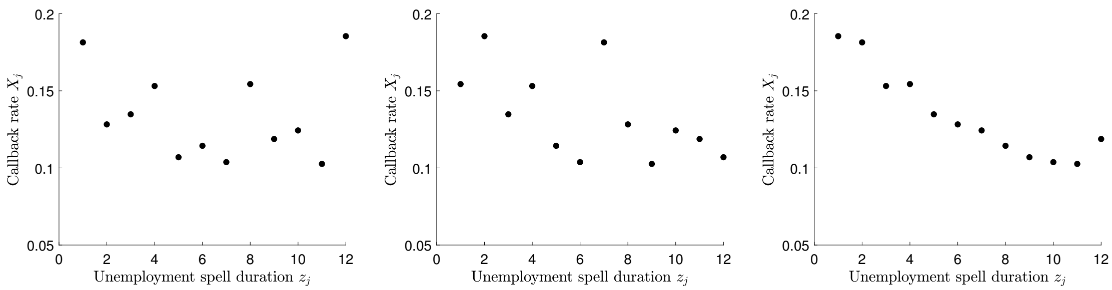
Figure 3 illustrates the conflict between the goals of communication and decision in a numerical example of this application. The top row of plots shows three example realizations of the data X that all correspond to the same rearranged decision $d^*(X)$, which is depicted in the bottom plot. The rearranged decision $d^*(X)$ dominates a decision of $d(X) = X$ in loss for any X for which $X_j > X_{j-1}$ for some j, and hence for the three realizations X depicted in the top row of plots. However, by pooling several different realizations of the data into the same decision, the rearranged decision $d^*(X)$ sacrifices information that is decision-relevant for some members of the audience.

Kroft et al. (2013, Figure 2, top panel) report the average callback rate as a function of months unemployed, which we may think of as analogous to reporting X, and which violates monotonicity in their data. Kroft et al. (2013, Figure 3) also report a sorted local linear regression estimate based on Chernozhukov et al (2009), which imposes monotonicity, and which we may think of as analogous to reporting $d^*(X)$. Eriksson and Rooth (2014, Table 5) report the callback rate as a function of months unemployed in the current spell, which violates monotonicity in their data. Eriksson and Rooth do not report any sorted or monotonicity-constrained estimates of this relationship.[10]

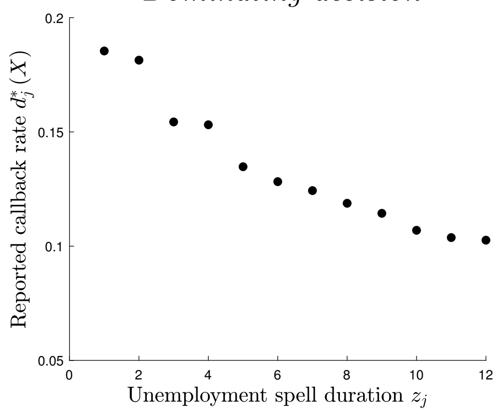
[10]Oberholzer-Gee (2008, Table 2) reports a linear probability model relating the callback probability to the number of months unemployed in the current spell, which respects monotonicity in his data among those currently unemployed.

Figure 3: Illustration of dominating decision in estimation of a monotone function

Example data realizations corresponding to the same dominating decision



Dominating decision



Note: The plots illustrate properties of the dominating decision in the estimation of a function that is known to be monotonically decreasing. The top row of plots depicts three possible realizations of the callback rates $X = (X_1, ..., X_J)$. These are the means of a binary outcome Y_i conditional on the unemployment duration $Z_i \in \mathcal{Z} = \{z_1, ..., z_J\} = \{1, ..., 12\}$. For each realization X, the decision $d = X$ is dominated in (quadratic) loss by the decision $d = d^*(X)$, which is depicted in the bottom plot. Under the dominating decision $d^*(X)$, all of the realizations X depicted in the top row of plots are reported identically, with the report depicted in the bottom plot.

The fact that both articles report estimates that violate monotonicity seems more consistent with the predictions of the communication model than with the predictions of the decision model. The fact that Kroft et al. (2013) additionally report sorted estimates seems consistent with the predictions of the decision model, perhaps suggesting a concern for both communication and decision-making in their context.

4.2 Optimal Treatment Assignment

Our second application is to optimal treatment assignment. Manski (2004) studies the problem of assigning one of a finite set of treatments to each member of some population, potentially as a function of a discrete covariate.

As a concrete example, consider the problem of an analyst who must make a clinical recommendation to an audience of physicians on the basis of the available evidence. This is a problem faced by many organizations, including medical learned societies and private publishers. Say that each physician's goal is to achieve the best average outcome for patients with each of a given set of attributes (e.g., diagnosis). We suppose these attributes are discrete, as in Manski (2004), and study the problem of recommending treatment to patients in a given attribute cell.

Formally, denote the available treatments (e.g., medications) by $t \in \{1, ..., T\}$ for $T \geq 2$. Suppose that $n \geq 1$ units (e.g., patients) are randomly allocated to each treatment t, and that for each unit i the analyst measures a binary outcome Y_i (e.g., an indicator for the resolution of symptoms). Let us further assume patient outcomes are independent, so it is without loss to represent the data for treatment t as a fraction of successes $X_t \in \left\{0, \frac{1}{n}, ..., 1\right\}$, with nX_t following a binomial distribution. The sample space is then

$$\mathcal{X} = \left\{0, \frac{1}{n}, ..., 1\right\}^T.$$

The unknown parameter is $(\theta_1, ..., \theta_T)$ where $\theta_t = \mathrm{E}[X_t]$ denotes the success probability for units assigned to treatment t. We assume that θ_t lies in a finite set $\Theta_0 \subset (0, 1)$ with $|\Theta_0| \geq 2$, so the parameter space is $\Theta = \Theta_0^T \subset (0, 1)^T$. We show in Online Appendix B.1.3 that our results for this application continue to hold for continuous parameter spaces $\Theta_0 = (0, 1)$ or $\Theta_0 = [0, 1]$.

The agent's (e.g., physician's) decision consists of picking a treatment or declining do so. Formally we take the decision space to be $\mathcal{D} = \{1, ..., T\} \cup \{\iota\}$ where ι corresponds to not picking a treatment. The agent's objective is to pick the best treatment which, following Manski (2004), we formalize by considering the regret loss

$$L(d, \theta) = \begin{cases} -\theta_d + \max_t \theta_t & \text{if } d \neq \iota \\ \max_t \theta_t & \text{if } d = \iota \end{cases}.$$

Declining to pick a treatment yields greater loss than picking any given treatment (e.g., because the patient cannot self-prescribe). The decision space in this case is too small to convey the full data, since $T + 1 = |\mathcal{D}| < |\mathcal{X}| = (n+1)^T$.

Classical decision-theoretic results for selection problems (Lehmann 1966, Eaton 1967) imply the following:

Claim 4. Consider the rule c^* that takes $c^*(X, V) = \arg\max_t X_t$ if the argmax is unique and otherwise randomizes uniformly over $\arg\max_t X_t$. This rule minimizes decision risk uniformly over $\Delta(\Theta)$ among rules that are invariant with respect to permutations of the treatments. Moreover, it is a Bayes decision rule for the agent a^* with $a^*(\theta) = \frac{1}{|\Theta|}$ for all $\theta \in \Theta$, and is minimax under the decision model.

The rule c^* is a special case of what Manski (2004) terms the "conditional empirical success" rule, and is related to the empirical welfare maximization procedures studied by Kitagawa and Tetenov (2018) and Athey and Wager (2019). The results of Stoye (2009) imply that c^* is a minimax decision rule for regret loss in the case of $T = 2$.

Claim 4 shows that the rule c^* is an appealing decision rule in several senses, and together with Theorem 1 implies that c^* is minimax under the communication model. The rule c^* is nevertheless inadmissible under the communication model. To see this, note first the following properties of the setting:

Claim 5. The decision $d = \iota$ is dominated in loss, and the effective size of the sample space is at least as large as the decision space, $N(\mathcal{X}) \geq |\mathcal{D}|$.

It follows by Proposition 3 that there is no rule $c \in \mathcal{C}$ that is admissible under both the decision model and the communication model, and no rule $c \in \mathcal{C}$ that is admissible under both the classical model and the communication model. Since, by Claim 4, c^* is a Bayes decision rule for the full-support prior $a^*(\theta) = \frac{1}{|\Theta|}$, it follows that c^* is admissible under the decision model and therefore, by Proposition 3 and Corollary 3(b), inadmissible, and not ω-optimal for any ω, under the communication model.

To understand why rule c^* is inadmissible under the communication model, note the following:

Claim 6. Under the communication model, the rule c^* is dominated by the rule $\tilde{c}$ that takes $\tilde{c}(X, V) = \iota$ if $\arg\max_t X_t = \{1, ..., T\}$ and $\tilde{c}(X, V) = c^*(X, V)$ otherwise.

When $\arg\max_t X_t = \{1, ..., T\}$, the data are uninformative about which treatment is best, in the sense that the likelihood for a given parameter value θ is unchanged if we permute which θ_t is associated to which treatment. The rule $\tilde{c}$ reflects this fact by reporting ι, while the rule c^* instead makes a guess at random. From the standpoint of decision-making, the rule c^* selects an undominated decision while the rule $\tilde{c}$ may not. From the standpoint of communication, the rule c^* obscures information that the rule $\tilde{c}$ does not. The information that c^* obscures is useful to agents whose priors are such that they wish to follow the empirical success rule only when the data are informative. Online Appendix B.7 extends the analysis to demonstrate a case in which the communication model favors reporting ι even when the data are informative, provided the amount of information in the data is small in comparison to the audience's beliefs.

In practice, analysts in situations like the one we have modeled sometimes express their ignorance rather than choosing a concrete recommendation at random. UpToDate is a private publisher that synthesizes medical research into clinical recommendations. As in the communication model, readers of these recommendations include practitioners who are free to make different clinical decisions. On the choice among selective serotonin reuptake inhibitors (SSRIs) to treat unipolar major depression in adults, UpToDate says "Given the lack of clear superiority in efficacy among antidepressants, selecting a drug is based on other factors, such as ... patient preference or expectations" (Simon 2019). Such a report seems more similar to $\tilde{c}$ than to c^*, and thus more consistent with the predictions of the communication model than with the predictions of the decision model.

5 Conclusions

We propose a model of scientific communication in which the analyst's report is designed to convey useful information to the agents in the audience, rather than, as in a classical model of statistics, to make a good decision or guess on these agents' behalf. We show conditions under which the proposed model predicts very different reporting rules from the classical model. We offer examples satisfying these conditions, and we argue that in practical situations similar to these examples, scientists' reports seem to be more consistent with the predictions of the proposed model than with the predictions of the classical model.

Proofs

This appendix provides proofs for the main results in the paper. Online Appendix A proves results for the Example and applications. Online Appendix B discusses extensions and additional results.

Proof of Proposition 1 We first prove the existence of an ω-optimal rule. Since the set of risk functions $\{\rho(c,\cdot) : c \in \mathcal{C}\}$ is compact in the supremum norm, the set of weighted average risks $\left\{\int_{\mathcal{A}} \rho(c,a)\, d\omega(a) : c \in \mathcal{C}\right\}$ is compact as well, and thus has a smallest element. This implies the existence of an ω-optimal rule.

We next show that any ω-optimal rule is admissible. Suppose that the result fails, so there exists some $c' \in \mathcal{C}$ that is ω-optimal but inadmissible. ω-optimality implies that

$$\int_{\mathcal{A}} \rho(c',a)\, d\omega(a) = \min_{c \in \mathcal{C}} \int_{\mathcal{A}} \rho(c,a)\, d\omega(a).$$

Inadmissibility implies the existence of $c'' \in \mathcal{C}$ such that $\rho(c',a) \geq \rho(c'',a)$ for all $a \in \mathcal{A}$ and $\rho(c',a') > \rho(c'',a')$ for some $a' \in \mathcal{A}$. By continuity of $\rho(c,a)$ in a, $\rho(c',a) > \rho(c'',a)$ 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,[11] and thus that

$$\int_{\mathcal{A}} \rho\left(c', a\right) d\omega\left(a\right) > \int_{\mathcal{A}} \rho\left(c'', a\right) d\omega\left(a\right).$$

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

Since we have shown that an ω-optimal rule exists for any ω, and that ω-optimal rules are admissible, it follows that an admissible rule exists.

We next prove the existence of minimax rules. To that end, note that $\mathcal{A}$ is a closed subset of a finite-dimensional simplex and so is compact. Since $\rho\left(c, a\right)$ is continuous in a and $\mathcal{A}$ is compact, $\sup_{a\in\mathcal{A}} \rho\left(c, a\right) = \max_{a\in\mathcal{A}} \rho\left(c, a\right)$ exists for all $c \in \mathcal{C}$. Moreover, since the set of risk functions $\left\{\rho\left(c, \cdot\right) : c \in \mathcal{C}\right\}$ is compact in the supremum norm, the set $\left\{\max_{a\in\mathcal{A}} \rho\left(c, a\right) : c \in \mathcal{C}\right\}$ is compact, and so has a minimum, which implies the existence of a minimax rule.

Finally, to prove the existence of an admissible minimax rule, define the set $\mathcal{C}_0 \subseteq \mathcal{C}$ of minimax rules and note that the set of minimax risk functions $\left\{\rho\left(c, \cdot\right) : c \in \mathcal{C}_0\right\}$ is a closed subset of a compact set, and thus is compact. Let us take a countable dense subset of $\mathcal{A}$, $\{a_1, a_2, ...\}$. For each $j \geq 1$, let $\mathcal{C}_j$ denote the subset of rules in $\mathcal{C}_{j-1}$ with minimal risk at a_j,

$$\mathcal{C}_j = \left\{c \in \mathcal{C}_{j-1} : \rho\left(c, a_j\right) = \min_{c' \in \mathcal{C}_{j-1}} \rho\left(c', a_j\right)\right\},$$

where the min on the right hand side is achieved by compactness of $\left\{\rho\left(c, \cdot\right) : c \in \mathcal{C}_{j-1}\right\}$. $\left\{\rho\left(c, \cdot\right) : c \in \mathcal{C}_j\right\}$ is a closed subset of a compact set, and so is again compact. Moreover, $\left\{\rho\left(c, \cdot\right) : c \in \mathcal{C}_j\right\}$ is non-empty by construction for all j, so $\cap_j \left\{\rho\left(c, \cdot\right) : c \in \mathcal{C}_j\right\}$ is non-empty by Cantor's Intersection Theorem. Any rule $c^* \in \cap_j \mathcal{C}_j \subseteq \mathcal{C}_0$ is minimax by definition but must also be admissible, since otherwise (by continuity of ρ in a) it would have been dropped at some finite step j. Thus, there exists an admissible minimax rule. □

Lemma 1. *Decision risk $R_a\left(c\right)$ and communication risk $R_a^*\left(c\right)$ are both continuous in a for all $c \in \mathcal{C}$. Moreover, the sets of risk functions $\left\{R_.\left(c\right) : c \in \mathcal{C}\right\}$ and*

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

$\{R^*_\cdot(c) : c \in \mathcal{C}\}$ *are compact in the supremum norm.*

Proof of Lemma 1 We first show continuity of decision risk. Note that the decision risk of a rule c can be written as

$$\sum_{\Theta} a(\theta)\,\mathrm{E}_\theta\left[L\left(c\left(X,V\right),\theta\right)\right]$$

where $L(d,\theta)$ is uniformly bounded. Hence, the decision risk is trivially continuous, and indeed Lipschitz, in a.

We next show continuity of communication risk. To this end, note that for any fixed $c \in \mathcal{C}$, if we define $\mathcal{B}$ to be the set of mappings $b : \mathcal{D} \times [0,1] \to \mathcal{D}$ then for any $c \in \mathcal{C}$ and any $a \in \Delta(\Theta)$ we have

$$R^*_a(c) = \min_{b\in\mathcal{B}} \mathrm{E}_a\left[L\left(b\left(c\left(X,V\right),V\right),\theta\right)\right] = \min_{b\in\mathcal{B}} R_a\left(b \circ c\right),$$

where the min is achieved by a binding rule $b_a \in \mathcal{B}$ with

$$b_a\left(c\left(X,V\right),V\right) \in \min_{d\in\mathcal{D}} \mathrm{E}_a\left[L\left(d,\theta\right)|c\left(X,V\right),V\right]$$

for all (X,V). However, the same argument used to establish continuity of decision risk implies that $\mathrm{E}_a\left[L\left(b\left(C,V\right),\theta\right)\right]$ is continuous in a. Indeed,

$$\begin{gathered}\sup_{c,b}\left|\mathrm{E}_a\left[L\left(b\left(c\left(X,V\right),V\right),\theta\right)\right] - \mathrm{E}_{\tilde{a}}\left[L\left(b\left(c\left(X,V\right),V\right),\theta\right)\right]\right| \leq \\ 2\max_{d,\theta}\left|L\left(d,\theta\right)\right|\textstyle\sum_{\Theta}\left|a\left(\theta\right)-\tilde{a}\left(\theta\right)\right|\end{gathered} \tag{4}$$

so $\mathrm{E}_a\left[L\left(b\left(c\left(X,V\right),V\right),\theta\right)\right]$ is Lipschitz in a (with respect to the L_1-norm) uniformly over b, c. This implies that $\min_{b\in\mathcal{B}} \mathrm{E}_a\left[L\left(b\left(c\left(X,V\right),V\right),\theta\right)\right]$ is Lipschitz in a as well, since for b_a the optimal rule for a, with

$$\min_{b\in\mathcal{B}} \mathrm{E}_a\left[L\left(b\left(c\left(X,V\right),V\right),\theta\right)\right] = \mathrm{E}_a\left[L\left(b_a\left(c\left(X,V\right),V\right),\theta\right)\right] = R^*_a(c),$$

we have

$$\mathrm{E}_a\left[L\left(b_{\tilde{a}}\left(c\left(X,V\right),V\right),\theta\right)\right] \geq \mathrm{E}_a\left[L\left(b_a\left(c\left(X,V\right),V\right),\theta\right)\right] = R^*_a(c)$$

$$\mathrm{E}_{\tilde{a}}\left[L\left(b_a\left(c\left(X,V\right),V\right),\theta\right)\right] \geq \mathrm{E}_{\tilde{a}}\left[L\left(b_{\tilde{a}}\left(c\left(X,V\right),V\right),\theta\right)\right] = R^*_{\tilde{a}}\left(c\right),$$

and thus

$$\left|R^*_a\left(c\right) - R^*_{\tilde{a}}\left(c\right)\right| \leq 2\max_{d,\theta}\left|L\left(d,\theta\right)\right| \sum_{\Theta}\left|a\left(\theta\right) - \tilde{a}\left(\theta\right)\right|.$$

Hence, communication risk is Lipschitz in a, with the same Lipschitz constant as decision risk.

To prove compactness, note that both the decision and communication risk functions are uniformly bounded by $\max_{d,\theta}\left|L\left(d,\theta\right)\right|$, and that the set of beliefs $\mathcal{A}$ is compact with respect to the L_1-norm $\sum_{\Theta}\left|a\left(\theta\right) - \tilde{a}\left(\theta\right)\right|$. Thus, the sets of decision and communication risk functions are bounded Lipschitz functions on a compact domain, and so are compact in the supremum norm by the Arzelà–Ascoli Theorem. □

Proof of Proposition 2 For admissibility, note that rule c dominates rule c' in the classical model if and only if

$$\mathrm{E}_\theta\left[L\left(c\left(X,V\right),\theta\right)\right] \leq \mathrm{E}_\theta\left[L\left(c'\left(X,V\right),\theta\right)\right] \text{ for all } \theta \in \Theta \tag{5}$$

with strict inequality for some θ. Note, however, that this implies

$$\mathrm{E}_a\left[L\left(c\left(X,V\right),\theta\right)\right] \leq \mathrm{E}_a\left[L\left(c'\left(X,V\right),\theta\right)\right]$$

for all a, with strict inequality for full-support a, and thus that c dominates c' in the decision model. Conversely, if c dominates c' in the decision model, then (5) is immediate. Moreover, there must be a strict inequality for some θ, or else the decision risk functions for c and c' would be the same.

For ω-optimality, for any full-support weights ω on $\Delta\left(\Theta\right)$, define

$$a_\omega\left(\theta\right) = \int_{\Delta(\Theta)} a\left(\theta\right) d\omega\left(a\right)$$

as the implied weight on Θ. Note that if ω has full support on $\Delta\left(\Theta\right)$ then a_ω has full support on Θ, and thus that if a rule c is optimal in the decision model with respect to the full-support weights ω, it is optimal in the classical model with respect to the full-support weights a_ω. Conversely, for any full-support weights a_ω in the classical model, let $\nu = \min_{\theta \in \Theta} a_\omega(\theta)$. Let η denote the uniform weight on $\Delta(\Theta)$, and let ω correspond to a mixture with weight $\nu|\Theta|$ on η and weight $1 - \nu|\Theta|$ on the degenerate weight which puts mass one on $\tilde{a}(\theta) = \frac{a_\omega(\theta) - \nu}{1 - \nu|\Theta|}$. By construction the weight ω has full support, and any rule that is optimal in the classical model with respect to a_ω is optimal in the decision model with respect to ω.

Finally note that the maximum risk in the decision model is the same as that in the classical model, in the sense that for all $c \in \mathcal{C}$,

$$\max_{a \in \mathcal{A}} \mathrm{E}_a \left[L\left(c\left(X, V\right), \theta\right)\right] = \max_{\theta \in \Theta} \mathrm{E}_\theta \left[L\left(c\left(X, V\right), \theta\right)\right].$$

Hence, equivalence of minimax rules in the two models is immediate. □

To prove several results that follow, it is helpful to restrict attention to the class of rules that are non-random conditional on the data and the public randomization device (that is, rules $c : \mathcal{X} \times [0,1] \to \mathcal{D}$, rather than $c : \mathcal{X} \times [0,1] \to \Delta(\mathcal{D})$). We can do so without increasing either decision or communication risk.

Lemma 2. *For any $c : \mathcal{X} \times [0,1] \to \Delta(\mathcal{D})$, there exists a publicly-randomized rule $\tilde{c} : \mathcal{X} \times [0,1] \to \mathcal{D}$ such that $R_a(\tilde{c}) = R_a(c)$ and $R^*_a(\tilde{c}) \le R^*_a(c)$, both for all $a \in \Delta(\Theta)$.*

Proof of Lemma 2 Note that we can construct a $U[0,1]$ random variable U independent of X and V such that for some function $\bar{c} : \mathcal{X} \times [0,1]^2 \to \mathcal{D}$, the conditional distribution of $\bar{c}(X, V, U)$ given (X, V) coincides with that of $c(X, V)$.[12]

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 $\tilde{c}(X, V) = \bar{c}(X, V_1, V_2)$, noting that $\tilde{c}(X, V)$ is non-random conditional on (X, V). The conditional distribution of $\tilde{c}(X, V)$ given X is the same as the conditional distribution of $c(X, V)$ given X for all X by construction. Hence, $\mathrm{E}_\theta\left[L\left(c\left(X, V\right), \theta\right)\right] = \mathrm{E}_\theta\left[L\left(\tilde{c}\left(X, V\right), \theta\right)\right]$ for all $\theta \in \Theta$, which implies that $R_a(c) = R_a(\tilde{c})$ for all $a \in \Delta(\Theta)$,

[12]For example, numbering the elements of $\mathcal{D}$ as $d_1, ..., d_{|\mathcal{D}|}$ and defining $p_j(X, V; c) = \Pr\{c(X, V) = d_j | X, V\}$, conditional on $c(X, V) = d_j$ let us take U to be uniformly distributed on $\left[\sum_{j' < j} p_{j'}(X, V; c), \sum_{j' \le j} p_{j'}(X, V; c)\right]$. 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 $\bar{c}$ so that $\bar{c}(X, V, U) = d_j$ if and only if $U \in \left[\sum_{j' < j} p_{j'}(X, V; c), \sum_{j' \le j} p_{j'}(X, V; c)\right]$, the conditional distribution of $\bar{c}(X, V, U)$ coincides with that of $c(X, V)$.

and so establishes the result for decision risk.

For communication risk, as in the proof of Lemma 1 above, we define $\mathcal{B}$ to be the set of mappings $b : \mathcal{D} \times [0,1] \to \mathcal{D}$ and let b_a denote the optimal rule for a. Let c denote the (potentially randomized) rule considered in the statement of the lemma, and b_a a corresponding binding rule. Note that by the definition of $\tilde{c}$

$$R_a^* (c) = \min_{b \in \mathcal{B}} R_a (b \circ c) = R_a (b_a \circ c) = R_a (b_a \circ \tilde{c}) \geq \min_{b \in \mathcal{B}} R_a (b \circ \tilde{c}) = R_a^* (\tilde{c}),$$

so $\tilde{c}$ has weakly lower communication risk than c.[13] $\square$

Proposition 3 extends to cases with general audiences $\mathcal{A} \neq \Delta(\Theta)$. To state the result for this more general case, we first need to define dominance in loss and the effective size of the sample space for a general audience $\mathcal{A}$.

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 $a' \in \mathcal{A}$.*

Definition 6. *Let $\mathcal{P}$ be the set of partitions of $\mathcal{X}$, with generic element $P \in \mathcal{P}$. Let $\mathcal{P}_{\mathcal{A}}^*$ denote the subset of $\mathcal{P}$ such that for every cell $\mathcal{X}_p \in P \in \mathcal{P}_{\mathcal{A}}^*$, each agent $a \in \mathcal{A}$ has at least one decision $d \in \mathcal{D}$ that is optimal for every $X \in \mathcal{X}_p$. That is,*

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

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

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

[13] To see why this inequality can be strict, it suffices to consider an example. Suppose that $\mathcal{D} = \{d_1, d_2\}$, $\mathcal{X} = \{x_1, x_2\}$, and that c implies $\Pr\{c(X, V) = d_j | V, X = x_j\} = 2/3$ for $j \in \{1, 2\}$. To achieve the same conditional distribution over $\tilde{c}(X, V) | X$ with a publicly-randomized rule $\tilde{c}$, it must be the case that $\mathcal{V}_{11} = \{v | \tilde{c}(x_1, v) = d_1\}$ and $\mathcal{V}_{22} = \{v | \tilde{c}(x_2, v) = d_2\}$ 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 $(\tilde{c}(X, V), V)$, whereas this is never possible based on observing $(c(X, V), V)$. Hence, for certain loss functions and priors, communication risk will be strictly smaller under $\tilde{c}$ than under c.

Proposition 4. *If (i) there exists a decision $d \in \mathcal{D}$ that is dominated in loss and (ii) $\mathcal{N}(\mathcal{X}, \mathcal{A}) \geq |\mathcal{D}|$, then any rule $c \in \mathcal{C}$ that is admissible under the decision model with audience $\mathcal{A}$ is inadmissible under the communication model with audience $\mathcal{A}$, and vice versa.*

Proof of Proposition 4 To prove this result, note first that under our full-support assumption on F_θ, any rule c that selects the dominated decision d with positive probability is dominated in decision risk by the rule c' that is equal to c except that it chooses the dominating decision d' whenever c chooses d. Hence, any rule c that selects d with positive probability is inadmissible under the decision model.

We next show that any rule c that selects d with probability zero is inadmissible under the communication model. By Lemma 2 we can limit attention to rules c that use only the public randomization device. Fix a given such rule c, and let $\mathcal{D}' \subset \mathcal{D}$ denote the set of decisions that c selects with positive probability. Since $|\mathcal{D}'| < |\mathcal{D}| \leq N(\mathcal{X}, \mathcal{A})$, we know that for every realization of the public randomization device V there is some $a_V \in \mathcal{A}$, some $\mathcal{X}_V \subseteq \mathcal{X}$, and some $d_V \in \mathcal{D}$ such that

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

and $c(X, V) = d_V$ for all $X \in \mathcal{X}_V$.[14]

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

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

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

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

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

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

$$\arg\min_{d \in \mathcal{D}} \mathrm{E}_{a^*}\left[L(d,\theta)\,|c(X,V) = d^*, V \in \mathcal{V}^*\right].$$

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

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

For $\bar{d}$ an element of $\mathcal{D}\backslash\mathcal{D}'$, let $\tilde{c}$ be the rule that reports $\tilde{c}(X,V) = \bar{d}$ when $V \in \mathcal{V}^*$ and $X = \tilde{X}$, and that agrees with c otherwise. Given $\tilde{c}(X,V)$, an agent can reconstruct $c(X,V)$, so the rule $\tilde{c}$ does not increase communication risk for any agent relative to c. For agent a^*, however,

$$\mathrm{E}_{a^*}\left[L\left(\tilde{d},\theta\right)|\tilde{c}(X,V) = \bar{d}, V \in \mathcal{V}^*\right] < \mathrm{E}_{a^*}\left[L\left(d^{**},\theta\right)|\tilde{c}(X,V) = \bar{d}, V \in \mathcal{V}^*\right],$$

so d^{**} is now sub-optimal for a^* conditional on observing $\left\{\tilde{c}(X,V) = \bar{d}, V \in \mathcal{V}^*\right\}$. This immediately implies that

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

$$\min_{d \in \mathcal{D}} \mathrm{E}_{a^*}\left[L(d,\theta)\,|c(X,V) = d^*, V \in \mathcal{V}^*\right].$$

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

Proof of Proposition 3 Follows immediately from Proposition 4 by setting $\mathcal{A} = \Delta(\Theta)$. $\square$

Proof of Corollary 1 Follows immediately from Propositions 2 and 3. $\square$

Proof of Corollary 3 Lemma 1 shows that the decision and communication models satisfy the conditions of Proposition 1. For part (a) of the corollary, note that since by Proposition 1 ω-optimality in the decision model implies admissibility in that model, under the conditions of the corollary it also implies inadmissibility in the communication model by Proposition 3. Moreover, since ω-optimality for communication implies admissibility for communication, the set of rules that are ω-optimal for decision and ω'-optimal for communication, where ω and ω' may be different, must not overlap.

For part (b) of the corollary, note that by Proposition 2, c is a Bayes decision rule with respect to a full-support prior if and only if it is ω-optimal in the decision model for some ω. The result thus follows by part (a).

Finally, for part (c) of the corollary, note that ω-optimality in the communication model implies admissibility for communication. By Proposition 3 this implies inadmissibility for decision, and thus by the argument in part (b) implies that the rule is not a Bayes decision rule with respect to a full-support prior. $\square$

Theorem 1 again extends to general (closed) audiences $\mathcal{A}$, though in this case we must also require that the audience be convex.

Theorem 2. *If the audience $\mathcal{A}$ is convex, any rule $c^* \in \mathcal{C}$ that is minimax under the decision model is minimax under the communication model.*

Proof of Theorem 2 Note that by continuity of decision risk in a (established in Lemma 1) and compactness of closed subsets of the finite-dimensional simplex, $\sup_{a\in\mathcal{A}} R_a(c)$ is achieved for all c. Note, next, that the set of decision risk functions $\{R_\cdot(c) : c \in \mathcal{C}\}$ is compact in the supremum norm by Lemma 1. Hence,

$$\inf_{c\in\mathcal{C}} \sup_{\mathcal{A}} R_a(c)$$

is achieved, so a min-max rule exists.

Thus, there exists $c^* \in \mathcal{C}$ with $\inf_{c\in\mathcal{C}} \max_{a\in\mathcal{A}} R_a(c) = \max_{a\in\mathcal{A}} R_a(c^*)$. Let $\tilde{c}^* : \mathcal{X} \to \Delta(\mathcal{D})$ be a rule that does not use the public randomization device such that the distribution of $\tilde{c}^*(X)|X$ is the same as that of $c^*(X, V)|X$ for all $X \in \mathcal{X}$. $R_a(\tilde{c}^*) = R_a(c^*)$ for all a, so $\max_{a\in\mathcal{A}} R_a(c^*) = \max_{a\in\mathcal{A}} R_a(\tilde{c}^*)$. For $\tilde{\mathcal{C}}$ the class of such rules, note that if we limit attention to $c \in \tilde{\mathcal{C}}$ we can cast the problem into the setting of Section 5 of Grünwald and Dawid (2004). Specifically, let us view $c \in \tilde{\mathcal{C}}$ 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 by assumption, the set of implied distributions for (X, θ) is closed and convex as well. Finally, the set of risk functions for $c : \mathcal{X} \to \mathcal{D}$ is finite, while the set of decision risk functions for $\tilde{\mathcal{C}}$ is the convex hull of a finite set and so is compact.

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

$$\max_{a \in \mathcal{A}} R_a(\tilde{c}^*) = R_{a^*}(\tilde{c}^*) = \min_{c \in \tilde{\mathcal{C}}} R_{a^*}(c). \tag{7}$$

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

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

so (7) and the fact that $R_a(\tilde{c}^*) = R_a(c^*)$ for all a implies that $\max_{a \in \mathcal{A}} R_a(c^*) = R_{a^*}(c^*) = \min_{c \in \mathcal{C}} R_{a^*}(c)$. Note, next, that by the definition of communication risk $\min_{c \in \mathcal{C}} R_{a^*}(c) = \min_{c \in \mathcal{C}} R^*_{a^*}(c)$. Thus, since $R_a(c) \geq R^*_a(c)$ for all $a \in \mathcal{A}$, $c \in \mathcal{C}$, we have $R_{a^*}(c^*) = R^*_{a^*}(c^*)$.

Combining this reasoning, we see that

$$\inf_{c \in \mathcal{C}} \max_{a \in \mathcal{A}} R_a(c) = \max_{a \in \mathcal{A}} R_a(c^*) = R_{a^*}(c^*) = R^*_{a^*}(c^*) = \min_{c \in \mathcal{C}} R^*_{a^*}(c).$$

Note, next, that by definition

$$\inf_{c \in \mathcal{C}} \max_{a \in \mathcal{A}} R_a(c) \geq \min_{c \in \mathcal{C}} \max_{a \in \mathcal{A}} R^*_a(c) \geq \min_{c \in \mathcal{C}} R^*_{a^*}(c).$$

Thus, since the first and last quantities are equal, we obtain that

$$\inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} R_a(c) = \inf_{c \in \mathcal{C}} R_{a^*}(c) = R_{a^*}(c^*) = R^*_{a^*}(c^*) = \inf_{c \in \mathcal{C}} R^*_{a^*}(c) = \inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} R^*_a(c),$$

as we wanted to show. □

Proof of Theorem 1 Immediate from Theorem 2, taking $\mathcal{A} = \Delta(\Theta)$. □

Proof of Corollary 4 Immediate from Proposition 3. □

References

Abadie, Alberto. Forthcoming. Statistical non-significance in empirical economics. *American Economic Review: Insights*.

Abadie, Alberto and Maximilian Kasy. 2019. Choosing among regularized estimators in empirical economics: The risk of machine learning. *Review of Economics and Statistics* 101(5): 743-762.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2017. Measuring the sensitivity of parameter estimates to estimation moments. *Quarterly Journal of Economics* 132(4): 1553-1592.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2020. Transparency in structural research. *NBER Working Paper No. 26631.*

Awasthi, S., V. K. Pande, and R. H. Fletcher. 2000. Effectiveness and cost-effectiveness of albendazole in improving nutritional status of pre-school children in urban slums. *Indian Pediatrics* 37(1): 19-29.

Athey, Susan and Stefan Wager. 2019. Efficient policy learning. *arXiv preprint* arXiv:1702.0289. Accessed on September 20 2019 at <https://arxiv.org/abs/1702.02896v5>.

Banerjee, Abhijit, Sylvain Chassang, Sergio Montero, and Erik Snowberg. Forthcoming. A theory of experimenters: robustness, randomization, and balance. *American Economic Review*.

Bergemann, Dirk and Stephen Morris. 2019. Information design: A unified perspective. *Journal of Economic Literature* 57(1): 44-95.

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.

Chen, Ying, Navin Kartik, and Joel Sobel. 2008. Selecting cheap-talk equilibria. *Econometrica* 76(1): 117-136.

Chernozhukov, Victor, Ivan Fernandez-Val, and Alfred Galichon. 2009. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* 96(3): 559-575.

Chetverikov, Denis, Andres Santos, and Azeem M. Shaikh. 2018. The econometrics of shape restrictions. *Annual Review of Economics* 10: 31-63.

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.

Croke, Kevin, Joan Hamory Hicks, Eric Hsu, Michael Kremer, and Edward Miguel. 2016. Does mass deworming affect child nutrition? Meta-analysis, cost-effectiveness, and statistical power. *World Bank Policy Research Working Paper No. 7921.*

Donnen, Philippe, Daniel Brasseur, Michèle Dramaix, Francoise Vertongen, Mweze Zihindula, Mbasha Muhamiriza, and Philippe Hennart. 1998. Vitamin A supplementation but not deworming improves growth of malnourished preschool children in eastern Zaire. *Journal of Nutrition* 128(8): 1320-1327.

Eaton, Morris L. 1967. Some optimum properties of ranking procedures. *The Annals of Mathematical Statistics* 38(1): 124-137.

Efron, Bradley. 1986. Why isn't everyone a Bayesian? *American Statistician* 40(1): 1-5.

Eriksson, Stefan and Dan-Olof Rooth. 2014. Do employers use unemployment as a sorting criterion when hiring? Evidence from a field experiment. *American Economic Review* 104(3): 1014-1039.

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.

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.

Jordan, Michael I., Jason D. Lee, and Yung Yang. 2018. Communication-efficient distributed statistical inference. *Journal of the American Statistical Association.*

Kamenica, Emir and Matthew Gentzkow. 2011. Bayesian persuasion. *American Economic Review* 101(6): 2590-2615.

Kitagawa, Toru and Aleksey Tetenov. 2018. Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica* 86(2): 591-616.

Kremer, Michael and Edward Miguel. 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72(1): 159-217.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo. 2013. Duration dependence and labor market conditions: Evidence from a field experiment. *The Quarterly Journal of Economics* 128(3): 1123–1167.

Kruger, Marita, Chad J. Badenhorst, and Erna P. G. Mansvelt. 1996. Effects of iron fortification in a school feeding scheme and anthelmintic therapy on the iron status and growth of six- to eight-year-old schoolchildren. *Food and Nutrition Bulletin* 17(1): 1-11.

Kwan, Yum K. 1999. Asymptotic Bayesian analysis based on a limited information estimator. *Journal of Econometrics* 88(1): 99-121.

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. 1966. On a theorem of Bahadur and Goodman. *The Annals of Mathematical Statistics* 37(1): 1-6.

Lehmann, E.L. and George Casella. 1998. *Theory of Point Estimation.* 2nd ed. New York: Springer-Verlag.

Manski, Charles F. 2004. Statistical treatment rules for heterogeneous populations.

Econometrica 72(4): 1221-1246.

Morris, Stephen. 1995. The common prior assumption in economic theory. *Economics and Philosophy* 11(2): 227-253.

Oberholzer-Gee, Felix. 2008. Nonemployment stigma as rational herding: A field experiment. *Journal of Economic Behavior & Organization* 65(1): 30-40.

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.

Simon, Gregory. 2019. Unipolar major depression in adults: choosing initial treatment. In D. Solomon (ed.), *UpToDate.* Accessed on September 19, 2019 at <https://www.uptodate.com/contents/unipolar-major-depression-in-adults-choosing-initial-treatment>.

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.

Spiess, Jann. 2018. Optimal estimation when researcher and social preferences are misaligned. *Harvard University Working Paper.* Accessed on December 27, 2019 at <https://scholar.harvard.edu/files/spiess/files/alignedestimation.pdf>.

Stoye, Jörg. 2009. Minimax regret treatment choice with finite samples. *Journal of Econometrics* 151(1): 70-81.

Stoye, Jörg. 2011. Statistical decisions under ambiguity. *Theory and Decision* 70(2): 129-148.

Stoye, Jörg. 2012. New perspectives on statistical decisions under ambiguity. *Annual Review of Economics* 4: 257-282.

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.

Wald, Abraham. 1950. *Statistical Decision Functions*. New York, NY: John Wiley & Sons.

Zhu, Yuancheng, and John Lafferty. 2018a. Distributed nonparametric regression under communication constraints. *Proceedings of the 35th International Conference on Machine Learning* 80.

Zhu, Yuancheng, and John Lafferty. 2018b. Quantized minimax estimation over Sobolev ellipsoids. *Information and Inference: A Journal of the IMA* 7(1): 31-82.

Zhang, Yuchen, John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. 2013. Information-theoretic lower bounds for distributed statistical estimation with communication constraints. *Advances in Neural Information Processing Systems* 26.

Online Appendix for

A Model of Scientific Communication

Isaiah Andrews, *Harvard University and NBER*
Jesse M. Shapiro, *Brown University and NBER*

February 2020

A Proofs for Claims in Example and Applications

Proof of Claim 1 For dominance in loss, note that since $0 \in \underline{\mathcal{X}}_M$ and $0 \in \overline{\mathcal{X}}_M$, $0 \in \mathcal{D}$ by construction. For any $d < 0$, $L(d,\theta) > L(0,\theta)$ for all $\theta \in \Theta$, so d is dominated in loss by $d' = 0 \in \mathcal{D}$.

To show that $N(\mathcal{X}) = |\underline{\mathcal{X}}_M| \times |\overline{\mathcal{X}}_M|$, note first that sufficiency of $(\underline{X}, \overline{X})$ implies that $N(\mathcal{X}) \leq |\underline{\mathcal{X}}_M| \times |\overline{\mathcal{X}}_M|$. To complete the proof we will show that for any $X = (\underline{X}, \overline{X}) \in \mathcal{X}$ and $X' = (\underline{X}', \overline{X}')$ such that $X \neq X'$, there exists an agent $a \in \Delta(\Theta)$ such that

$$\arg\min_{d \in \mathcal{D}} \mathrm{E}_a [L(d,\theta)|X] \cap \arg\min_{d \in \mathcal{D}} \mathrm{E}_a [L(d,\theta)|X'] = \emptyset, \tag{8}$$

which in turn implies that $N(\mathcal{X}) \geq |\underline{\mathcal{X}}_M| \times |\overline{\mathcal{X}}_M|$.

Note that under the assumed form for the distribution F_0, the joint likelihood of the control group outcomes satisfies the monotone likelihood ratio property in $\underline{X}$.[15] In particular, for any agent with a non-degenerate prior on $\underline{\theta}$, the posterior distribution on $\underline{\theta}$ is strictly increasing in $\underline{X}$ in the sense of first-order stochastic dominance. Fix the prior on $\overline{\theta}$ as the prior which puts mass one on the largest value in Θ_0, and consider

[15] Specifically, the joint density is given by $\prod_{i=1}^{\underline{n}} \exp(\underline{\theta} X_i) h(\underline{\theta}) g(X_i) = \exp(\underline{\theta} \cdot \underline{n} \cdot \underline{X}) h(\underline{\theta})^{\underline{n}} \prod_{i=1}^{\underline{n}} g(X_i)$, and for $\underline{X}' > \underline{X}$ and $\underline{\theta}' > \underline{\theta}$,

$$\frac{\prod_{i=1}^{\underline{n}} \exp(\underline{\theta}' X_i') h(\underline{\theta}') g(X_i')}{\prod_{i=1}^{\underline{n}} \exp(\underline{\theta} X_i') h(\underline{\theta}) g(X_i')} > \frac{\prod_{i=1}^{\underline{n}} \exp(\underline{\theta}' X_i) h(\underline{\theta}') g(X_i)}{\prod_{i=1}^{\underline{n}} \exp(\underline{\theta} X_i) h(\underline{\theta}) g(X_i)}.$$

the class of all priors on $\underline{\theta}$, $\Delta(\Theta_0)$, noting that since we have taken $\overline{\theta}$ as large as possible all of these priors respect the bounds on the parameter space.

Since the data have full support under all $\theta \in \Theta$, the implied class of posteriors for θ conditional on $\left(\underline{X}, \overline{X}\right)$ is the same as the class of priors, and $\overline{X}$ does not matter for the posterior. The posterior risk function for agent a conditional on X is

$$\sum_{\theta \in \Theta} (d-\theta)^2 \, a(\theta|X).$$

Strict convexity of the squared error loss implies that each agent's set of optimal actions conditional on $\underline{X}$ contains at most two elements. Let us fix a value $\underline{X} < \max\{\underline{\mathcal{X}}_M\}$ and consider an agent a who, conditional on $\underline{X}$, is indifferent between two decisions d and d' with $d < d'$ (and prefers these to all other decisions). Existence of such an agent follows from our assumptions. Specifically, since $\underline{n} \geq 3$, and $\max_{\overline{\theta} \in \Theta_0} \mathrm{E}_{F_0(\overline{\theta})}[X_i] - \min_{\underline{\theta} \in \Theta_0} \mathrm{E}_{F_0(\underline{\theta})}[X_i] > 2/3$, we know that there exist $n_1 < n_2 \leq \underline{n}$ such that $\frac{1}{n_1}, \frac{1}{n_2} \in (0, \max_{\theta \in \Theta_0} ATE(\theta))$. However, upper and lower bounds on $ATE(\theta)$ after fixing $\overline{\theta}$ at its upper bound are $\max_{\theta \in \Theta_0} ATE(\theta)$ and 0, respectively. Thus, by the definition of $\mathcal{D}$ there exist d and d' in $\mathcal{D}$ with $d, d' \in (0, \max_{\theta \in \Theta_0} ATE(\theta))$, from which the claim follows.

Note that since agent a is indifferent between d and d', we must have $\max_{\overline{\theta} \in \Theta_0} \mathrm{E}_{F_0(\overline{\theta})}[X_i] - \mathrm{E}_a[\underline{\theta}|X] = \frac{1}{2}d + \frac{1}{2}d'$. Since $|\Theta_0| \geq 2$, we know that there exists $\underline{\theta}' \in \Theta_0$ with $\underline{\theta}' \neq \mathrm{E}_a[\underline{\theta}|X]$. Assume $\underline{\theta}' < \mathrm{E}_a[\underline{\theta}|X]$ (if not, we can use the same argument with $\underline{\theta}' > \mathrm{E}_a[\underline{\theta}|X]$). If we define a_ε as the agent with $\varepsilon > 0$ more prior mass on $\underline{\theta}'$, and ε less prior mass on all other parameter values, then for ε sufficiently small a_ε strictly prefers d' to any other action conditional on X, but strictly prefers d to d' conditional on any $\underline{X}' > \underline{X}$. Since we can repeat this argument for all $\underline{X} < \max\{\underline{\mathcal{X}}_M\}$, this verifies (8) for all X, X' such that $\underline{X}' \neq \underline{X}$. However, we can likewise repeat the same argument to verify (8) for all X, X' such that $\overline{X}' \neq \overline{X}$, which proves the claim. □

Proof of Corollary 2 Follows immediately from Claim 1, Proposition 3, and Corollary 1. □

Proof of Claim 2 Note that for all $\theta \in \Theta$, θ_j is decreasing in j by assumption. We can consider d and θ as step functions defined on the interval $[0, J]$ with steps at the integers, and $\|d-\theta\|_2$ is just the L_2 distance in this case. For d^* which takes $d_j^* = d_{(j)}$, Proposition 1 of Chernozhukov et al (2009) shows that for any $d \in \mathcal{D}$ with $d_j > d_{j-1}$ for some j and any $\theta \in \Theta$,

$$\|d-\theta\|_2 \geq \|d^*-\theta\|_2$$

with strict inequality if $\theta_j < \theta_{j-1}$. Hence

$$\|d-\theta\|_2^2 \geq \|d^*-\theta\|_2^2,$$

for all $\theta \in \Theta$, with strict inequality for some some $\theta' \in \Theta$, and d is dominated in loss. □

Proof of Claim 3 To prove the claim, we need to show that any two distinct realizations of the data, say X and X', induce distinct optimal actions for some agent, in the sense that there exists $a \in \Delta(\Theta)$ with

$$\arg\min_{d\in\mathcal{D}} \mathrm{E}_a\left[L(d,\theta)\,|X\right] \cap \arg\min_{d\in\mathcal{D}} \mathrm{E}_a\left[L(d,\theta)\,|X'\right] = \emptyset.$$

To this end, suppose without loss of generality that for some j, $X_j \neq X_j'$. Let us consider the set of agents $\mathcal{A}_j$ who put probability one on $\theta_{j'} = \max\{\Theta_0\}$ for $j' < j$, and probability one on $\theta_{j'} = \min\{\Theta_0\}$ for $j' > j$. Hence, for these agents monotonicity imposes no restrictions on θ_j and, moreover, the only decision-relevant data is X_j. However, the binomial distribution has exponential family structure with $f(x;t) = \exp(tx)\,h(t)\,g(x)$. Hence, the same argument used to prove Claim 1 in the running example establishes the result. Since we can repeat this argument for all X and X' with $X \neq X'$, this proves the claim. □

Proof of Claim 4 Theorems 4.2 and 4.3 of Eaton (1967) establish the invariant optimality and minimaxity claims in the classical model. Since we have taken the audience to consist of the set of all priors, Proposition 2 shows that minimaxity holds in the classical model if and only if it holds in the decision model, and the same is true of invariant optimality. For the Bayes part of the claim, note that the prior a^* implies that θ_t and θ_s are independent for all $s \neq t$, so $\mathrm{E}_{a^*}[\theta_t|X]$ is a function of X_t alone. Moreover, by the monotone likelihood ratio assumption $\mathrm{E}_{a^*}[\theta_t|X_t]$ is strictly increasing in X_t. Hence, $\arg\max_t \mathrm{E}_{a^*}[\theta_t|X] = \arg\max_t X_t$, which proves the claim. $\square$

Proof of Claim 5 For the first part of the claim, note that $d = 1$ yields a strictly lower loss than choosing $d = \iota$ for all $\theta \in \Theta$.

For the second part of the claim, note that it suffices to show that it holds for a restricted version of the audience, $\tilde{\mathcal{A}} \subseteq \Delta(\Theta)$. Specifically, let us consider the audience consisting of only three agents, a_0, a_1, and a_2.

Agent a_0 has a uniform prior on Θ, $a_0(\theta) = a^*(\theta) = \frac{1}{|\Theta|}$. This again implies that θ_t is independent of θ_s for all $s \neq t$. By the monotone likelihood ratio property, provided $\arg\max_t X_t$ is unique, this agent strictly prefers to set $d = \arg\max_t X_t$. When $\arg\max_t X_t$ is non unique, by contrast, this agent strictly prefers $d \in \arg\max_t X_t$ to $d \notin \arg\max_t X_t$, but is indifferent among $d \in \arg\max_t X_t$.

Note, next, that since

$$a(\theta|X) = \frac{f(X;\theta)\,a(\theta)}{\sum_\Theta f\left(X;\tilde{\theta}\right) a\left(\tilde{\theta}\right)},\ \mathrm{E}_a[\theta|X] = \sum_\Theta \theta a(\theta|X).$$

for $f(X;\theta)$ the probability mass function of F_θ, and F_θ has full support for all $\theta \in \Theta$, $\mathrm{E}_a[\theta|X]$ is continuous in a. Hence, there exists an open neighborhood $\mathcal{N}(a_0)$ around a_0 such that all agents $a \in \mathcal{N}(a_0)$ strictly prefer to set $d \in \arg\max_t X_t$ to $d \notin \arg\max_t X_t$ for all realizations of X. Within this neighborhood, there is an agent a_1 who when

$$\arg\max_t X_t = \{1, ..., T\}$$

strictly prefers $d = 1$, and an agent a_2 who strictly prefers $d = 2$ conditional on the same event. This immediately implies, however, that $N(\mathcal{X}, \mathcal{A}) \geq T + 1$, since

$$\left(\arg\min_{d\in\mathcal{D}} \mathrm{E}_{a_0}[L(d,\theta)|X], \arg\min_{d\in\mathcal{D}} \mathrm{E}_{a_1}[L(d,\theta)|X], \arg\min_{d\in\mathcal{D}} \mathrm{E}_{a_2}[L(d,\theta)|X]\right)$$

$$= \begin{cases} (\arg\max_t X_t, \arg\max_t X_t, \arg\max_t X_t) & \text{when } \arg\max_t X_t \text{ is a singleton} \\ (\arg\max_t X_t, 1, 2) & \text{when } \arg\max_t X_t = \{1, ..., T\} \end{cases},$$

where the right hand side takes $T + 1$ distinct values. $\square$

Proof of Claim 6 Since agents are free to randomize conditional on observing $\tilde{c}(X,V) = \iota$, $\tilde{c}$ yields weakly smaller communication risk for all agents than c^*. To show that there is a strict inequality for some agents, consider agents a for whom $\mathrm{E}_a\left[\theta_t | \arg\max_t X_t = \{1,...,T\}\right]$ is non-constant across t, while for all (X,V) $\arg\max_t \mathrm{E}_a\left[\theta_t | c^*(X,V)\right] = c^*(X,V)$. Note in particular that the agent with a uniform prior $a^*(\theta) = 1/|\Theta|$ has $\arg\max_t \mathrm{E}_{a^*}\left[\theta_t | c^*(X,V)\right] = c^*(X,V)$, so such agents a exist by the continuity of conditional expectations in a. When $\arg\max_t X_t = \{1,...,T\}$, the decision taken by these agents is uniformly randomized under the rule c^*, while under the rule $\tilde{c}$ they are able to pick a decision they strictly prefer to uniform randomization. Since these agents are still free to set $d = \tilde{c}(X,V)$ when $\tilde{c}(X,V) \neq \iota$, this establishes that they have strictly lower communication risk under $\tilde{c}$ than under c^*. □

B Additional Results Referenced in the Text

B.1 Continuous Versions of Example and Applications

This section extends the recurring deworming example, along with the monotone estimation and treatment assignment applications discussed in Section 4, to continuous settings. For the deworming example and monotone estimation application we show that changing from decision to communication risk can reverse dominance orderings in a setting with Θ, $\mathcal{X}$, and $\mathcal{D}$ all continuous, while for the treatment assignment application we show that the results discussed in the main text extend unchanged to the case with continuous Θ but discrete $\mathcal{X}$ and $\mathcal{D}$. We also extend our main result on admissibility, Proposition 3, to the case of continuous Θ.

B.1.1 Continuous Version of Deworming Example

Consider a Gaussian variant of the discrete deworming example developed in the main text. Specifically, suppose that the data consist of control- and treatment-group means from a randomized trial. Let us take $\mathcal{X} = \mathbb{R}^2$, and assume that the two means are independent and normally distributed with known variances $\underline{\sigma}^2, \bar{\sigma}^2 > 0$,

$$\begin{pmatrix} \underline{X} \\ \bar{X} \end{pmatrix} \sim N\left(\begin{pmatrix} \underline{\theta} \\ \bar{\theta} \end{pmatrix}, \begin{pmatrix} \underline{\sigma}^2 & 0 \\ 0 & \bar{\sigma}^2 \end{pmatrix} \right).$$

Normality with known variance can be motivated by the asymptotic normality of sample means (along with consistent estimability of their asymptotic variance).

Note that the average treatment effect is simply the difference between the treatment and control means, $ATE(\theta) = \overline{\theta} - \underline{\theta}$, and that the sample average treatment effect $\overline{X} - \underline{X}$ is again an unbiased estimator for $ATE(\theta)$. We again restrict the parameter space such that the average treatment effect is non-negative, considering

$$\Theta = \left\{\theta \in \mathbb{R}^2 : \overline{\theta} - \underline{\theta} \geq 0\right\},$$

and take the decision space to correspond to the support of the treatment-control difference, $\mathcal{D} = \mathbb{R}$. We consider quadratic loss $L(d, \theta) = (d - ATE(\theta))^2$, and take the audience to consist of all possible priors $\Delta(\Theta)$. For any rule $c : \mathcal{X} \times [0,1] \to \Delta(\mathcal{D})$ and any prior $a \in \Delta(\Theta)$, we define decision and communication risk as in the discrete case, with the difference that, unlike in the discrete case, decision and communication risk may now be infinite for some rules and priors.

To highlight the distinction between decision and communication risk, we consider two particular rules. First, the treatment-control difference,

$$c'(X, V) = c'(X) = \overline{X} - \underline{X}$$

and, second, the treatment-control difference censored below at zero,

$$c''(X, V) = c''(X) = \max\left\{\overline{X} - \underline{X}, 0\right\}.$$

Since all audience members agree that the average treatment effect is non-negative, and the distribution of $c'(X)$ has support equal to $\mathbb{R}$, $c''(X)$ has strictly lower decision risk than c' for all a,

$$R_a(c'') < R_a(c') \text{ for all } a \in \Delta(\Theta).$$

At the same time, c'' censors the data. Note in particular that distinct realizations of $c'(X)$ induce distinct posterior expectations for all agents a with non-degenerate priors on $ATE(\theta)$, $Var_a(ATE(\theta)) > 0$. Hence, by strict convexity of the loss, censoring negative estimates strictly increases communication risk for these agents, while not reducing it for any other agent. Thus, we see that $R_a^*(c'') \geq R_a^*(c')$ for all $a \in \Delta(\Theta)$, with strict inequality when $Var_a(ATE(\theta)) > 0$. Hence, switching from decision to communication risk reverses the dominance ordering between $c'(X)$ and $c''(X)$. The rule $c'(X)$ preferred by the communication model again appears closer to what is reported in practice.

B.1.2 Continuous Version of Monotone Estimation Application

Next, let us consider a continuous version of the monotone estimation application discussed in the main text. Specifically, suppose we observe a normal random vector $X \in \mathcal{X} = \mathbb{R}^J$ with independent elements $X_j \sim N(\theta_j, \sigma_j^2)$, and σ_j strictly positive and known for all j. As with the previous extension, this may again be motivated by asymptotic normality of sample means.

Suppose that $\theta \in \Theta_0$ for each j, for Θ_0 some interval with a nonempty interior (e.g. $\Theta_0 = \mathbb{R}$), and that θ_j is known to be decreasing in j so the joint parameter space is

$$\Theta = \left\{\theta \in \Theta_0^J : \theta_1 \geq \theta_2 \geq ... \geq \theta_J\right\}.$$

We take $\mathcal{D} = \mathbb{R}^J$, consider the squared error loss,

$$L(d, \theta) = \|d - \theta\|_2^2 = \sum_j (d_j - \theta_j)^2,$$

and again take the audience to consist of all possible priors, $\Delta(\Theta)$.

As in the deworming example above we compare two decision rules. The first reports the raw data

$$c'(X, V) = c'(X) = X,$$

while the second reports sorted estimates

$$c''(X, V) = c''(X) = d^*(X),$$

where as in the main text $d^*(d)$ sorts the elements of d in decreasing order, so $d_1^*(d) = \max_j d_j$ and $d_J^*(d) = \min_j d_j$.

The results of Chernozhukov et al (2009) again show that c' dominates c'' in decision risk, and in particular that $R_a(c'') \leq R_a(c')$ for all $a \in \Delta(\Theta)$, while $R_a(c'') < R_a(c')$ for a such that $\Pr_a\{\theta_j < \theta_{j-1} \text{ for some } j\} > 0$.

At the same time, $c''(X)$ is a transformation of $c'(X)$, so

$$R_a^*(c'') \geq R_a^*(c') \text{ for all } a \in \Delta(\Theta).$$

Note, next, that any two distinct realizations of $c'(X)$ induce distinct posterior expectations for some agent. To see that this is the case, suppose that X and $\tilde{X}$ differ in their jth component, and consider an agent $\tilde{a}$ with degenerate priors on $\theta_{j'}$ for $j' \neq j$ and a non-degenerate normal prior on θ_j (truncated to obey the monotonicity restrictions). If $\tilde{a}$ observes $c'(X)$ then their posterior mean $\mathrm{E}_{\tilde{a}}[\theta_j|c'(X)] = \mathrm{E}_{\tilde{a}}[\theta_j|X_j]$ is strictly monotonic in X_j. If this agent instead observes $c''(X)$, any value of $c''(X)$ with at least two distinct elements is consistent with multiple values of X_j, which implies that

$$\Pr{}_{\tilde{a}}\{\mathrm{E}_{\tilde{a}}[\theta_j|c''(X)] \neq \mathrm{E}_{\tilde{a}}[\theta_j|c'(X)]\} > 0.$$

Jensen's inequality thus implies that $R_{\tilde{a}}^*(c'') > R_{\tilde{a}}^*(c')$. Thus, we again see that switching from decision risk to communication risk reverses the dominance ordering between c' and c''.

B.1.3 Continuous Version of Treatment Choice Application

Define $\mathcal{X}$, $\mathcal{D}$, and $L(d, \theta)$ as in Section 4.2, but take $\Theta = (0,1)^T$ (or $\Theta = [0,1]^T$) and again consider the audience $\mathcal{A} = \Delta(\Theta)$. The results of Lehmann (1966) and Eaton (1967) continue to apply, and imply that the rule $c^*(X, V)$ discussed in the main text uniformly minimizes decision risk among all decision rules invariant with respect to permutation of the treatments, and is minimax. Moreover, c^* is a Bayes decision rule for the agent a^* with an independent $U[0,1]$ prior on each θ_t, and is admissible in the classical model by Theorem 4.3 of Eaton (1967). Admissibility in the decision model then follows immediately by the same argument as in the proof of Proposition 2.

At the same time, the rule $\tilde{c}$ defined in the main text continues to dominate c^* in the communication model. To see that this is the case, note that given $\tilde{c}(X, V)$ the agent is again free to randomize uniformly when $\tilde{c}(X, V) = \iota$, which yields a random decision with the same conditional distribution a $c^*(X, V)|X$ for all X. Thus, $R_a^*(\tilde{c}) \leq R_a^*(c^*)$ for all $a \in \Delta(\Theta)$. To see that this inequality is strict for some $a \in \Delta(\Theta)$, consider the restricted set of priors $\Delta\left(\Theta_0^T\right)$ for Θ_0 finite, $|\Theta_0| \geq 2$. Claim 6 implies that there is a strict inequality for some $a \in \Delta\left(\Theta_0^T\right)$.

B.1.4 Admissibility Conflict with Continuous Parameter Space

A corollary of Proposition 4 is that the conclusions of Proposition 3 can be extended to some cases with continuous Θ.

Corollary 5. *Suppose that for an infinite (and potentially uncountable) set Θ but finite $\mathcal{X}$ and $\mathcal{D}$, there exists a decision $d \in \mathcal{D}$ that is dominated in loss, in the sense that there exists $d' \in \mathcal{D}$ with $L(d,\theta) \geq L(d',\theta)$ for all $\theta \in \Theta$ with strict inequality for at least one $\theta \in \Theta$. Suppose further that for some finite subset $\tilde{\Theta} \subset \Theta$, $\mathcal{N}\left(\mathcal{X}, \Delta\left(\tilde{\Theta}\right)\right) \geq |\mathcal{D}|$. Then any rule c that is admissible in decision risk is inadmissible in communication risk and vice versa.*

The dominance in loss condition holds for continuous Θ in the deworming example, monotone estimation application, and treatment choice application. Thus, the admissibility conflict extends to these three settings as well under conditions on Θ analogous to those in the discrete case. (Specifically, we maintain that $\sup_{\bar{\theta}\in\Theta_0} \mathrm{E}_{F_0(\bar{\theta})}[X_i] - \inf_{\underline{\theta}\in\Theta_0} \mathrm{E}_{F_0(\underline{\theta})}[X_i] > 2/3$ in the deworming example, and $\sup\{\Theta_0\} - \inf\{\Theta_0\} > 2/3$ in the monotone estimation application.)

Proof of Corollary 5 Note first that any rule with $\Pr_\theta\{c(X,V) = d\} > 0$ for some $\theta \in \Theta$ is inadmissible in decision risk by the same argument as in the proof of Proposition 4. Next, consider any rule such that $\Pr_a\{c(X,V) = d\} = 0$ for all $a \in \Delta\left(\tilde{\Theta}\right)$. Let us construct a rule $\tilde{c}$ as in the proof of Proposition 4, and note that by construction $R_a^*(\tilde{c}) \leq R_a^*(c)$ for all $a \in \Delta(\Theta)$, with a strict inequality for some $a \in \Delta\left(\tilde{\Theta}\right)$. Thus, c is inadmissible in the communication model, as we wanted to show. $\square$

B.2 Heterogeneous Loss Functions and Likelihoods

In Section 2, we claimed that models with heterogeneous loss functions and likelihoods can be cast into our setting. To see that this is the case, suppose we begin with a model with an audience of agents $\mathcal{A}$ where agent a's prior is $\pi_a \in \Delta(\Theta)$, their likelihood function is $F_{a,\theta}$, and their loss function is $L_a(d,\theta)$. Here we do not identify agent a with their prior since agents differ on multiple dimensions. Let us further suppose that the loss is uniformly bounded above

$$\sup_{a\in\mathcal{A},d\in\mathcal{D},\theta\in\Theta} L_a(d,\theta) < \infty,$$

and that the probability mass function $f_a(x;\theta)$ of $F_{a,\theta}$ is uniformly bounded away from zero,

$$\inf_{a\in\mathcal{A},x\in\mathcal{X},\theta\in\Theta} f_a(x;\theta) \geq \eta > 0.$$

Under the assumption of uniformly bounded loss, for $M = |\Theta| \times |\mathcal{D}|$ we can express

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

where $L_0(d,\theta) = 0$, while for $m \geq 1$ the functions $L_m(d,\theta)$ are proportional to $1\left\{d = d_{j(m)}\right\} 1\left\{\theta = \theta_{k(m)}\right\}$, $\left\{\left(d_{j(m)},\theta_{k(m)}\right) : m = 1,...,M\right\} = \mathcal{D}\times\Theta$, l_a satisfies $l_a(m) \geq 0$ for all m, and $\sum_{m=0}^{M} l_a(m) = 1$. Intuitively, $l_a(m) = L_a\left(d_{j(m)},\theta_{k(m)}\right)/L_m\left(d_{j(m)},\theta_{k(m)}\right)$, so for two agents a and a', $l_a(m)/l_{a'}(m)$ measures how much a loses, relative to a', from the decision-parameter pair $\left(d_{j(m)},\theta_{k(m)}\right)$.

Likewise, we can express

$$\Pr_{F_{a,\theta}}\{X = x\} = \sum_{n=1}^{N} p_a(n) f_n(x;\theta)$$

where

$$f_n(x;\theta) = \begin{cases} 1 - \eta(|\mathcal{X}| - 1) & \text{if } \theta = \theta_{k(n)} \text{ and } x = x_{r(n)} \\ \eta & \text{otherwise} \end{cases},$$

$\left\{\left(\theta_{k(n)},x_{r(n)}\right) : n = 1,...,N\right\} = \Theta\times\mathcal{X}$, $p_a(n)$ satisfies $p_a(n) \geq 0$ for all n, and $\sum_{n=0}^{N} p_a(n) = 1$.

Using these observations, we can cast this example into our baseline setting by augmenting the parameter space. Specifically, define

$$\Theta^* = \Theta \times \{0,...,M\} \times \{1,...,N\}.$$

For each agent a, define a's prior on this augmented space as $\pi_a^*(\theta^*) = \pi_a(\theta) l_a(m) p_a(n)$, and let agents share the homogeneous loss function

$$L\left(d,\theta^{*}\right)=L_{m}\left(d,\theta\right),$$

and the homogeneous likelihood

$$F_{\theta^{*}}=F_{n,\theta},$$

where $F_{n,\theta}$ has mass function $f_{n}\left(\cdot;\theta\right)$. Since the prior imposes mutual independence between (θ,m,n), agent $a's$ posterior risk from action d conditional on (X,V) is

$$\mathrm{E}_{a}\left[L\left(d,\theta^{*}\right)|X,V\right]=\frac{\sum_{\theta\in\Theta}\sum_{n=1}^{N}\sum_{m=0}^{M}L_{m}\left(d,\theta\right)f_{n}\left(X;\theta\right)p_{a}\left(n\right)l_{a}\left(m\right)\pi_{a}\left(\theta\right)}{\sum_{\theta\in\Theta}\sum_{n=1}^{N}\sum_{m=0}^{M}f_{n}\left(X;\theta\right)p_{a}\left(n\right)l_{a}\left(m\right)\pi_{a}\left(\theta\right)}=$$

$$\frac{\sum_{\theta\in\Theta}\sum_{n=1}^{N}f_{n}\left(X;\theta\right)p_{a}\left(n\right)\sum_{m=0}^{M}L_{m}\left(d,\theta\right)l_{a}\left(m\right)\pi_{a}\left(\theta\right)}{\sum_{\theta\in\Theta}\sum_{n=1}^{N}f_{n}\left(X;\theta\right)p_{a}\left(n\right)\pi_{a}\left(\theta\right)}=$$

$$\frac{\sum_{\theta\in\Theta}L_{a}\left(d,\theta\right)f_{a}\left(x;\theta\right)\pi\left(\theta\right)}{\sum_{\theta\in\Theta}f_{a}\left(x;\theta\right)\pi\left(\theta\right)}=\mathrm{E}_{a}\left[L_{a}\left(d,\theta\right)|X,V\right]$$

and so coincides with the posterior risk in the initial heterogeneous prior, heterogeneous loss, heterogeneous likelihood model. Since this holds for all (X,V), the posterior risk conditional on $\left(c\left(X,V\right),V\right)$ likewise coincides for all possible c. Hence, all risk calculations likewise coincide, and we have successfully recast the initial heterogeneous prior, heterogeneous loss, heterogeneous likelihood model as a heterogeneous prior, homogeneous loss, homogeneous likelihood model with a particular audience $\mathcal{A}\subset\Delta\left(\Theta^{*}\right)$. Whether the class of priors $\{\pi_{a}^{*}:a\in\mathcal{A}\}$ is closed and/or convex will depend on the original priors π_{a}, losses L_{a}, and likelihoods $F_{a,\theta}$.

In conjunction with heterogeneous loss functions and likelihoods, our approach can also allow heterogeneous decision spaces $\mathcal{D}_{a}$ and parameter spaces Θ_{a} provided $|\mathcal{D}_{a}|=|\mathcal{D}|$ and $|\Theta_{a}|=|\Theta|$ for all $a\in\mathcal{A}$. To see that this is the case, note that for each a we can define a bijection ϕ_{a} between $\mathcal{D}_{a}$ and $\mathcal{D}$, and a bijection ψ_{a} between Θ_{a} and Θ. Using these bijections we can again regard the loss and likelihood for agent a as being defined on $\mathcal{D}$ and Θ, and use the same argument as above.

B.3 Regret-based Payoff Functions

For a given loss function $L : \mathcal{D} \times \Theta \to \mathbb{R}_{\geq 0}$, consider the regret loss function $\tilde{L} : \mathcal{D} \times \Theta \to \mathbb{R}_{\geq 0}$ formed by subtracting the minimized loss:

$$\tilde{L}(d, \theta) = L(d, \theta) - \min_{d' \in \mathcal{D}} L(d', \theta).$$

Regret loss functions of this kind have been considered by e.g. Manski (2004). Defining the communication, decision, and frequentist risk functions with respect to the regret loss leaves all of the results in Section 3 unchanged (since the regret loss function is simply a different choice of loss, and our results do not depend on the form of the loss).

For a given risk function $\rho : \mathcal{C} \times \mathcal{A} \to \mathbb{R}_{\geq 0}$, we may alternatively consider the regret risk function $\tilde{\rho} : \mathcal{C} \times \mathcal{A} \to \mathbb{R}_{\geq 0}$ formed by subtracting the best-possible risk:

$$\tilde{\rho}(c, a) = \rho(c, a) - \inf_{c' \in \mathcal{C}} \rho(c', a).$$

Regret risk functions of this kind are considered in Stoye (2012). Defining the regret communication, decision, and frequentist risk functions leaves the results on admissibility and $\omega-$optimality in Sections 3.1 and 3.2 unchanged (since the set of ω-optimal and admissible rules are the same in both cases). We do not know whether the results on minimaxity in Section 3.3 extend to this case.

B.4 Selecting Natural Permutations

Let ψ permute the elements of $\mathcal{D}$, and let $\psi \circ c : \mathcal{X} \times [0, 1] \to \Delta(\mathcal{D})$ be the rule with realization $\psi(c(X, V))$. It is immediate that $(\psi \circ c) \in \mathcal{C}$ and that $R_a^*(\psi \circ c) = R_a^*(c)$. By contrast, in general $R_a(\psi \circ c) \neq R_a(c)$. That is, communication risk is invariant to permutations, whereas decision risk is not. As a result, the communication model has the unrealistic implication that if, for example, $\mathcal{D}$ is a subset of $\mathbb{R}$ and is symmetric around zero, all agents are indifferent between the report c and the report $-c$.

It is possible to modify the communication model to eliminate this unrealistic implication. Define for any rule $c \in \mathcal{C}$ the set of rules

$$\mathcal{Q}(c) = \{c' \in \mathcal{C} : R_a^*(c') = R_a^*(c) \, \forall a \in \mathcal{A}\}$$

that are equivalent in communication risk to the rule c. Now say that a rule $c^* \in \mathcal{C}$ is

- *naturally admissible in communication risk* if c^* is admissible in communication risk and there exists no rule $c \in \mathcal{Q}(c^*)$ such that $R_a(c) \leq R_a(c^*)$ for all $a \in \mathcal{A}$, with strict inequality for at least one $a \in \mathcal{A}$.

- *naturally ω-optimal in communication risk* if c^* is ω-optimal in communication risk and
$$\int_{\mathcal{A}} R_a(c^*)\, d\omega(a) = \inf_{c \in \mathcal{Q}(c^*)} \int_{\mathcal{A}} R_a(c)\, d\omega(a).$$

- *naturally minimax in communication risk* if c^* is minimax in communication risk and
$$\sup_{a \in \mathcal{A}} R_a(c^*) = \inf_{c \in \mathcal{Q}(c^*)} \sup_{a \in \mathcal{A}} R_a(c).$$

For each optimality criterion, naturalness selects from among the communication-optimal rules those that are not worse in decision risk than some other rule that is equivalent in communication risk. The resulting rules are "natural" in the sense that they will not lead to unnecessary losses if their reports are taken literally as recommended decisions. This approach is similar in spirit to the refinement of cheap-talk equilibria that arises from an arbitrarily small cost of lying (Chen et al. 2008).

The resulting refinement does not affect the relationship between the communication and decision models.

Proposition 5. *There exists a rule c^* that is admissible in decision risk and naturally admissible in communication risk if and only if there exists a rule $\tilde{c} \in \mathcal{C}$ that is admissible in both decision and communication risk. The same holds for ω-optimality and minimaxity.*

Proof of Proposition 5 We first consider admissibility. Since a rule is naturally admissible in communication risk only if it is admissible in communication risk, the "only if" part of the statement is trivial. For the "if" part, note that if c^* is admissible in both decision risk and communication risk, then it is admissible in communication risk, and admissible in decision risk relative to the restricted class of rules $\mathcal{Q}(c^*)$. However, this is the definition of natural admissibility. The same argument works for ω-optimality and minimaxity. □

B.5 Derivation of Quadratic Loss in Deworming Example

Let $k \geq \max\{\mathcal{D}\}$ be the private marginal cost of the medication with $k-d$ the market price with a subsidy (or tax) $d \in \mathcal{D}$. Assume that households' private willingness-to-pay is distributed uniformly on $[0, W]$ for $W \geq k - \min\{\mathcal{D}\}$, so that the share of households purchasing the medication at market price $k-d$ is given by

$$\frac{W-(k-d)}{W} \in [0,1].$$

Each purchase has a social benefit not internalized by households given by $\tilde{\theta} = ATE(\theta)$ (e.g. because malnutrition has an impact on children's long-run outcomes that is not fully internalized by parents). The government's payoff is given by the sum of total surplus from the purchases, plus the social benefit:

$$\left(\frac{W-(k-d)}{W}\right)\left(\frac{W+(k-d)}{2} - \left(k-\tilde{\theta}\right)\right)$$

where $\frac{W+(k-d)}{2}$ is the average private willingness-to-pay conditional on purchase and $(k - ATE(\theta))$ is the social marginal cost. We can rewrite the government's payoff as

$$\frac{1}{2W}\left(W^2 - 2W\left(k-\tilde{\theta}\right) - \left(d-\tilde{\theta}\right)^2 + \left(k-\tilde{\theta}\right)^2\right).$$

Subtracting the best possible payoff (which arises when $d = \tilde{\theta}$) gives

$$-\frac{1}{2W}\left(d-\tilde{\theta}\right)^2$$

which is directly proportional to the assumed loss.

B.6 Complete Class Theorems for Decision and Communication Models

In this section we establish complete class theorems for decision and communication risk. To state this result, we allow a general closed audience $\mathcal{A} \subseteq \Delta(\Theta)$, and say that a rule $c^* \in \mathcal{C}$ is $\underline{\omega}$-optimal in decision risk if and only if

$$\int R_a(c^*)\, d\underline{\omega}(a) = \inf_{c\in\mathcal{C}} \int R_a(c)\, d\underline{\omega}(a),$$

where unlike for ω-optimality the support of the probability measure $\underline{\omega}$ may be a strict subset of $\mathcal{A}$. We define $\underline{\omega}$-optimality in communication risk analogously.

Proposition 6. *A rule c is admissible in decision risk for the audience $\mathcal{A}$ only if it is $\underline{\omega}$-optimal in decision risk for some $\underline{\omega}$. Likewise, a rule is admissible in communication risk only if it is $\underline{\omega}$-optimal in communication risk for some $\underline{\omega}$.*

For decision risk, this result is an immediate consequence of the classical complete class theorem. By contrast, communication risk is nonlinear in a, and so requires a slightly different argument (and in particular depends on the public randomization device).

Proof of Proposition 6 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 communication risk. Note first that the presence of V ensures that the set of communication risk functions is convex. Specifically, for any pair of rules c_1, $c_2 \in \mathcal{C}$ 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 $(c_\alpha (X, V), V)$,

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

We show continuity of communication risk in Lemma 1, 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 communication risk. □

B.7 Extension of Optimal Treatment Assignment Example

This section extends the analysis of optimal treatment assignment in Section 4.2 of the main text to show that when agents have sufficiently informative beliefs, it may be communication-preferred to report ι even in some cases without exact ties. To develop these results we consider a restricted audience $\bar{\mathcal{A}} \subset \Delta(\Theta)$.

Claim 7. Suppose that for a closed audience $\bar{\mathcal{A}}$ and some non-empty set $\mathcal{E} \subseteq \mathcal{X}$,

$$\arg\max_t \mathrm{E}_a[\theta_t|X] = \arg\max_t \mathrm{E}_a[\theta_t] \text{ for all } a \in \bar{\mathcal{A}}, X \in \mathcal{E}. \tag{9}$$

Then the rule $\bar{c}$ which takes $\bar{c}(X,V) = c^*(X,V)$ when $X \notin \mathcal{E}$ and $\bar{c}(X,V) = \iota$ when $X \in \mathcal{E}$ has weakly lower communication risk for all $a \in \bar{\mathcal{A}}$ than does the rule c^*.

Claim 8. If in addition to the conditions of Claim 7,

$$\left\{X : \arg\max_t X_t = \{1, ..., T\}\right\} \cap \mathcal{E} \neq \emptyset, \tag{10}$$

there exists $a \in \bar{\mathcal{A}}$ with $\arg\max_t \mathrm{E}_a[\theta_t|c^*(X,V), V] = c^*(X,V)$ for all (X,V), and $\arg\max_t \mathrm{E}_a[\theta_t]$ a singleton, then $\bar{c}$ dominates c^* in communication risk.

Claim 9. Finally, if the conditions of Claim 7 hold, the conditions of Claim 8 hold for all $a \in \bar{\mathcal{A}}$, and the audience $\bar{\mathcal{A}}$ is invariant under permutation of the treatments (that is, the set of priors is unchanged by relabeling the treatments), then c^* is minimax in decision risk but not communication risk for the audience $\bar{\mathcal{A}}$.

Condition (9) in Claim 7 formalizes what it means for $X \in \mathcal{E}$ to be uninformative for the audience $\bar{\mathcal{A}}$: no agent in this audience changes their action in response to observing $X \in \mathcal{E}$. Note that $\mathcal{E}$ and $\bar{\mathcal{A}}$ are tightly linked here. For example, if we take $\bar{\mathcal{A}}$ to contain only agents with dogmatic priors, then all data is uninformative and we can take $\mathcal{E} = \mathcal{X}$. By contrast, if we take $\bar{\mathcal{A}} = \{a^*\}$ to consist of the agent with an uninformative prior discussed in Claim 4, then the largest possible $\mathcal{E}$ is

$$\mathcal{E} = \left\{X : \arg\max_t X_t = \{1, ..., T\}\right\}.$$

Claim 8 is intuitive as well. The condition $\arg\max_t \mathrm{E}_a[\theta_t|c^*(X,V), V] = c^*(X,V)$ means that agent a has a sufficiently uninformative prior that they are willing to follow the recommendation of the minimax rule. By contrast, the fact that $\arg\max_t \mathrm{E}_a[\theta_t]$ is a singleton means this agent's prior is sufficiently informative that, absent any data, they have a unique preferred decision. Putting these two conditions together, however, means that this agent would strictly prefer to avoid the randomization induced by the rule c^*.

Finally, for Claim 9, note that, since the audience is invariant, minimaxity of c^* follows from the fact that it minimizes risk in the class of invariant decision rules. However, the conditions of Claim 9 imply that the audience of $\bar{\mathcal{A}}$ is not convex, since the convex hull of any invariant audience necessarily contains a prior $\bar{a}$ such that

$$\arg\max_t \mathrm{E}_{\bar{a}}\left[\theta_t\right] = \{1, ..., T\}$$

which we have ruled out by assumption. Hence, Claim 9 illustrates that with a non-convex audience, minimax decision rules need not be minimax communication rules.

Proof of Claim 7 Note that all agents have the option to choose $d \in \arg\max_t \mathrm{E}_a\left[\theta_t\right]$ conditional on observing $\bar{c}\left(X,V\right) = \iota$, while choosing $d \in \arg\max_t \mathrm{E}_a\left[\theta_t|\bar{c}\left(X,V\right),V\right]$ conditional on observing $\bar{c}\left(X,V\right) \neq \iota$. By the definition of $\mathcal{E}$ this yields a weakly lower expected loss for agent a than choosing some $d \in \arg\max_t \mathrm{E}_a\left[\theta_t|c^*\left(X,V\right),V\right]$. $\square$

Proof of Claim 8 If $\arg\max_t \mathrm{E}_a\left[\theta_t\right]$ is a singleton for a given agent a and (9) holds, then conditional on $X \in \mathcal{E}$ agent a strictly prefers not to randomize their decision. At the same time, since $\arg\max_t \mathrm{E}_a\left[\theta_t|c^*\left(X,V\right)\right] = c^*\left(X,V\right)$, under the rule c^* this agent's decision is random conditional on the data when

$$X \in \mathcal{E} \cap \left\{X : \arg\max_t X_t = \{1, ..., T\}\right\}.$$

As above, since the agent is free to choose $d = \bar{c}\left(X,V\right)$ conditional on $\bar{c}\left(X,V\right) \neq \iota$ and $d = \arg\max_t \mathrm{E}_a\left[\theta_t\right]$ conditional on $\bar{c}\left(X,V\right) = \iota$, we see that $\bar{c}$ yields a strictly lower communication risk for this agent. Since we have shown in the proof of Claim 7 that $\bar{c}$ yields weakly lower communication risk than c^* for all $a \in \bar{\mathcal{A}}$, $\bar{c}$ dominates c^*. $\square$

Proof of Claim 9 Since the set $\bar{\mathcal{A}}$ is a closed subset of the simplex and communication risk is continuous in a, the proof of Claim 8 implies that

$$\inf_{a\in\bar{\mathcal{A}}}\left\{R_a^*\left(c^*\right)-R_a^*\left(\bar{c}\right)\right\}>0.$$

Hence, the maximum communication risk of c^* for the audience $\bar{\mathcal{A}}$ is strictly larger than that of $\bar{c}$, which proves the second part of the claim. To prove the first part of the claim, note that by the assumed invariance of the audience $\bar{\mathcal{A}}$, the maximum decision risk of a rule c is bounded below by the risk of the rule $\tilde{\psi}\circ c$ where $\tilde{\psi}$ randomly permutes the treatments. Note, however, that the rule $\tilde{\psi}\circ c$ is invariant to relabeling of the treatments by construction. The results of Lehmann (1966) imply, however, that $R_a\left(c^I\right)\geq R_a\left(c^*\right)$ for any invariant rule c^I and any $a\in\mathcal{A}$. Thus, we have that for any rule c

$$\sup_{a\in\bar{\mathcal{A}}}R_a\left(c\right)\geq\sup_{a\in\bar{\mathcal{A}}}R_a\left(\tilde{\psi}\circ c\right)\geq\sup_{a\in\bar{\mathcal{A}}}R_a\left(c^*\right),$$

which proves that c^* is minimax in decision risk. $\square$