A Model of Scientific Communication

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

March 2021

Abstract

We propose a positive 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 report directly determines the payoff. We identify settings in which the predictions of the proposed model differ from those of the classical model, and seem to better match practice.

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 Grants No. 1654234 and 1949047, the Sloan Research Fellowship, the Silverman (1968) Family Career Development Chair at MIT, and the Eastman Professorship and Population Studies and Training Center at Brown University. 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, the Eitan Berglas School of Economics, the National Bureau of Economic Research, and the Wharton School, and from comments by conference discussants Jann Spiess and James Heckman. 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 optimality in econometrics.[1] The classical theory of point estimation, for instance, 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 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, positive model of empirical science to capture scientific situations of this kind. In the proposed *communication model*, defined in Section 2, the analyst makes a report to an audience based on some data. After observing the analyst's report, but not the underlying data, each agent in the audience takes their optimal decision. Agents differ in their priors or loss functions, and may therefore have different optimal decisions following a given report. A reporting rule (specifying a distribution of reports for each realization of the data) induces an expected loss for each agent, which we call the rule's *communication risk.*

We compare the proposed communication model with a *decision model* in which the analyst selects a decision that directly determines the loss for all agents. The *decision risk* of a rule for a given agent is then the expected loss under the agent's prior from taking the decision prescribed by the rule.[2] The decision model generalizes the classical frequentist model, and the decision model's implications coincide with those

[1] See Lehmann and Casella (1998) for a textbook treatment of statistical decision theory and Stoye (2012) and Manski (2019) for recent discussions of its relation to econometrics.

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

of the classical model in a particular sense. By contrast, we find that the implications of the decision model can be very different from those of the communication model.

Section 3 presents an example in which the communication and decision models imply opposite dominance orderings of the same rules. In the example, the 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. Although deworming medication is known to (weakly) improve nutrition, sampling error means that the treatment-control difference may be negative. Under 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). Thus, an uncensored rule dominates a censored one under the communication model, while the reverse is true under the decision model. We claim, and illustrate by example, that a scientist choosing a report for a research article would be unlikely to censor. We also develop some general properties of the communication model that are suggested by the example.

Section 4 presents an example in which the communication and decision models disagree in an even stronger sense. In this example, the analyst conducts a randomized controlled trial to determine, from a finite set of options, the optimal treatment for a medical condition. When all of the treatments show equally promising effects in the trial, the decision model implies that it is optimal for the analyst to randomize among the treatments. By contrast, under the communication model, randomization discards the information that the treatments showed similar effects, which is useful to an agent who has a prior or preference in favor of one of them. Thus, a rule that reports that the trial was inconclusive dominates one that randomizes among the treatments under the communication model, while the reverse is true under the decision model. In fact, we show that *any* rule that is undominated (admissible) under the decision model in this example must be dominated (inadmissible) under the communication model, and vice versa. Again, we illustrate by example that the implications of the communication model seem to better match practice in at least some situations, and we develop some general results suggested by the example in an appendix.

Section 5 looks beyond dominance comparisons to consider alternative ways of selecting rules. One is to minimize weighted average risk which, under the decision model, corresponds to selecting Bayes decision rules. If all agents receive positive weight, then (under regularity conditions) weighted average risk inherits any ordering implied by dominance, and the conflicts in the preceding examples stand. Another way to select rules is to minimize the maximum risk over agents in the audience. Here we find more agreement between the two models in the sense that if the class of beliefs in the audience is convex, then (under regularity conditions) 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.

We illustrate both results in an example, based on GMM estimation, in which an analyst needs to combine multiple potentially misspecified moment conditions to learn about a structural parameter of interest. We characterize, respectively, rules that minimize weighted average decision risk and communication risk, and show how and why they differ. We further derive minimax decision rules, show that they are not minimax optimal for communication when the audience is non-convex, and discuss why they become minimax optimal for communication when the audience is convex.

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 given the data. When agents are instead heterogeneous, the distinction can be consequential, because different agents may prefer different decisions (or estimates).

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] Our

[3]See also Efron (1986) and Poirier (1988). A related literature (e.g., Pratt 1965; Kwan 1999; Abadie 2020; Abadie and Kasy 2019; Frankel and Kasy forthcoming) assesses the Bayesian interpretation of frequentist inference. Another literature (e.g., Zhang et al. 2013; Jordan et al. 2018; Zhu and Lafferty 2018) considers the problem of distributing statistical estimation and inference across

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. Andrews et al. (2020) study the implications of communication risk for structural estimation in economics (see also Andrews et al. 2017).

Our setting is also related to the literature on comparisons of experiments following Blackwell (1951, 1953), reviewed, for example, in Le Cam (1996) and Torgersen (1991). 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. We formalize the connection to sufficiency, which plays an important role in this literature, in Section 3.3.

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, which Spiess (2018), Banerjee et al. (2020), and others have recently considered in a statistical context.

2 Model

An analyst observes data $X \in \mathcal{X}$, for $\mathcal{X}$ a sample space. The distribution of X is governed by the parameter $\theta \in \Theta$, $X \sim F_\theta$, for Θ a parameter space. The analyst publicly commits to a rule $c : \mathcal{X} \to \Delta(\mathcal{S})$ that maps from realizations of the data X to a distribution over reports $s \in \mathcal{S}$, for $\mathcal{S}$ a signal space and $\Delta(\mathcal{S})$ the set of distributions on $\mathcal{S}$. Let $\mathcal{C}$ denote the set of all such rules, and with a slight abuse of notation let $c(X) \in \mathcal{S}$ denote the realization from a given rule $c \in \mathcal{C}$.

The analyst's report $c(X)$ is transmitted to a set of agents indexed by a. Each

multiple machines when communication is costly. Brown (1975) considers a setting with a collection of possible loss functions, while the literature on robust Bayesian decision theory (see e.g. Gilboa and Schmeidler 1989; Stoye 2012) analyzes decision rules with respect to classes of priors.

agent a is identified with a prior $a \in \Delta(\Theta)$ on the parameter space. We will call the set $\mathcal{A} \subseteq \Delta(\Theta)$ of such priors the *audience.* While we interpret the audience as a collection of agents, our model can be interpreted as one in which there is a single agent who possesses additional information unavailable to the analyst.[4]

After receiving the analyst's report $c(X)$ each agent a takes a decision $d \in \mathcal{D} \subseteq \mathcal{S}$, for $\mathcal{D}$ a decision space. It will sometimes be useful to focus on rules whose reports are valid decisions, i.e. rules $c : \mathcal{X} \to \Delta(\mathcal{D})$. We term such rules *decision rules* and let $\mathcal{B}$ denote the set of all such rules, where since $\mathcal{D} \subseteq \mathcal{S}$, we have $\mathcal{B} \subseteq \mathcal{C}$.

After taking the decision d, the agent a realizes the loss $L(d, \theta) \geq 0$. The analyst is benevolent and wishes to minimize the *ex ante* expected loss, or *risk,* of each agent under the agent's own prior. We consider two notions of risk. The first, which we call *decision risk*, is the expected loss to the agent from following the decision recommended by the analyst's report. Formally, for $c \in \mathcal{B}$, the decision risk $R_a(c)$ is

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

where $\mathrm{E}_a[\cdot]$ denotes the expectation under a's prior. The second notion of risk, which we call the *communication risk*, is the expected loss when each agent updates their beliefs based on the analyst's report and then selects a decision that is optimal under their updated beliefs. Formally, for $c \in \mathcal{C}$ the communication risk $R_a^*(c)$ is

$$R_a^*(c) = \mathrm{E}_a\left[\inf_{d \in \mathcal{D}} \mathrm{E}_a\left[L(d, \theta) \,|c(X)\right]\right].$$

For given audience $\mathcal{A}$ and loss $L(\cdot, \cdot)$, we will call the model with rules $\mathcal{B}$ and risk functions $R_a(\cdot)$ the *decision model*, and the model with rules $\mathcal{C}$ and risk functions $R_a^*(\cdot)$ the *communication model.* The assumption that all agents share a common loss function is without loss of generality, as a model with heterogeneous loss functions can always be re-parameterized as one with a homogeneous loss and a richer parameter θ.

Both the decision model and the communication model evaluate the expected loss with respect to the agent's own prior. The key difference between the decision model

[4]Under this interpretation, $\mathcal{A}$ is the set of posterior beliefs that the agent may hold after receiving the additional information.

and the communication model is that, under the decision model, the expected loss is evaluated as if the agent is forced to adopt the decision recommended by the analyst's report, whereas under the communication model, the expected loss is evaluated as if each agent takes their optimal decision conditional on the analyst's report.

If we take the audience $\mathcal{A}$ to be the set of point-mass priors on Θ, i.e., the vertices of $\Delta(\Theta)$, then the decision risk is the frequentist risk (Lehmann and Casella 1998, equation 1.10), and the decision model coincides with the classical model. If we instead take the audience $\mathcal{A}$ to be the set of all possible priors on Θ, i.e., $\Delta(\Theta)$, then the decision model still selects the same rules as the classical model under many standard optimality criteria (see Stoye 2012 for discussion). We therefore focus on comparing the decision and communication models.

The implications of the decision and communication models coincide if we take the audience $\mathcal{A}$ to be a singleton with unique element a^*. In this case, under the decision model the analyst will choose a rule c^* such that $c^*(X)$ minimizes $\mathrm{E}_{a^*}[L(d,\theta)|X]$ almost surely. Any such rule is also optimal under the communication model. If $\mathcal{A}$ instead contains multiple priors this logic need not apply and, as we show below, the two models can have quite different implications.

Interpretation of the Decision and Loss

We pause to highlight two ways to interpret the decision $d \in \mathcal{D}$ and 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 their 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 observing 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.

3 Conflict in Dominance Ordering

We will say that a rule c dominates another rule c' under a given model if the rule c achieves weakly lower risk for all agents in the audience and strictly lower risk for some. In this section, we show by example that the decision and communication models can imply opposite dominance orderings, in the sense that c dominates c' in the communication model but c' dominates c in the decision model.

3.1 A Treatment Effect with a Sign Constraint

An analyst observes data on weight gain for a sample of children enrolled in a randomized trial of deworming drugs (anthelmintic therapy). For the N_C children in the control group, weight gain X_i is distributed as $X_i \sim N(\theta_C, \sigma^2)$. For the N_T children in the treatment group, weight gain X_i is distributed as $X_i \sim N(\theta_T, \sigma^2)$. Thus the sample space is $\mathcal{X} = \mathbb{R}^{N_C+N_T}$. We assume that weight gain is independent across children so that the control group mean $\overline{X}_C$ and treatment group mean $\overline{X}_T$ follow

$$\begin{pmatrix} \overline{X}_C \\ \overline{X}_T \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_C \\ \theta_T \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{N_C} & 0 \\ 0 & \frac{\sigma^2}{N_T} \end{pmatrix} \right).$$

The variance σ^2 and group sizes (N_C, N_T) are commonly known. The average treatment effect of deworming drugs on child weight is $\theta_T - \theta_C$. Suppose that this effect is known *a priori* to be nonnegative, and in particular $\Theta = \{(\theta_C, \theta_T) \in \mathbb{R}^2 : \theta_T \geq \theta_C\}$.

The audience consists of governments who must decide how much to subsidize (or tax) deworming drugs. The governments face a loss $L(d, \theta) = (d - (\theta_T - \theta_C))^2$ for d the per-unit subsidy, with $d < 0$ denoting a tax. The set of feasible decisions is $\mathcal{D} = \mathbb{R}$. We assume that the audience $\mathcal{A}$ consists of the set of all distributions such that $\theta_T - \theta_C$ is a zero-truncated normal. All statements in this section continue to apply when $\mathcal{A} = \Delta(\Theta)$.

Consider two decision rules, c and c', defined as

$$c(X) = \overline{X}_T - \overline{X}_C, \;\; c'(X) = \max\{c(X), 0\}.$$

The rule c reports the difference in means between the treatment and control groups. The rule c' censors this report at 0.

Claim 1. *Rule c' dominates rule c under the decision model. Rule c dominates rule c' under the communication model.*

Proofs are collected in Appendix A, but we sketch the argument here. Start with the decision model. Because all governments accept that $\theta_T \geq \theta_C$, a tax on deworming drugs is never optimal. Yet, the rule c will sometimes recommend a tax. Under the decision model, such a recommendation incurs an unnecessarily large loss, because it is worse than recommending a neutral policy $d = 0$.

Next consider the communication model. Although all governments accept that $\theta_T \geq \theta_C$, in cases where $\overline{X}_T - \overline{X}_C < 0$ the realized value of $\overline{X}_T - \overline{X}_C$ is nevertheless informative about the true value of $\theta_T - \theta_C$. Intuitively, the lower is $\overline{X}_T - \overline{X}_C$, the stronger is the evidence for a small value of $\theta_T - \theta_C$. The rule c preserves this information, whereas the rule c' discards it. Even though every government's optimal subsidy d is nonnegative, there is no benefit to the censoring in c', because each government can simply censor its own decision d based on the information conveyed by c.

We can compare the implications of the decision and communication models to observed practice in a situation similar to the example. Kruger et al. (1996) conducted an early randomized controlled trial of the effect of deworming drugs on children's growth. A separate randomization was used to study the effect of iron-fortified soup. 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 nonnegative, 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) did not publish a censored estimate, nor did any of the four studies that Croke et al. (2016) identify as implying negative point estimates of the effect of deworming drugs on weight.[5]

3.2 Discussion

We have focused on a scenario where the audience consists of policymakers, so the loss captures the value of setting the right policy. We may alternatively envision the loss as capturing the scientific community's desire for a good guess of the true average

[5]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).

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).

We have focused on rules that have range $\mathcal{D}$ and are therefore decision rules. This is natural under the decision model but is restrictive under the communication model. To illustrate, suppose that $\mathcal{S}$ contains $\mathbb{R}^2$ and consider the rule c'' with

$$c''(X) = \left(\overline{X}_C, \overline{X}_T\right).$$

Claim 2. *(i) The rule c'' dominates the rule c under the communication model. (ii) The rule c'' achieves weakly lower risk for all agents than does any other rule under the communication model.*

Because the rule c'' conveys more information than the rule c, it dominates the rule c under the communication model. Moreover, because the statistic $c''(X)$ is sufficient for θ, rule c'' is weakly better than any other rule for any agent under the communication model. Interestingly, Kruger et al. (1996) report group means for the control and treatment groups, and do not explicitly report the difference $\overline{X}_T - \overline{X}_C$.

We have also focused on a situation in which the tension between the decision and communication models arises due to an *a priori* constraint on the parameter space. While illustrative, that is not the only situation in which the tension arises. Imagine, for example, that the randomized controlled trial is run at $J \geq 3$ sites j, each of which is associated with its own parameters $\theta_j = \left(\theta_C^j, \theta_T^j\right)$. We drop the sign constraints, so that $\Theta = \mathbb{R}^{2J}$, and take $\mathcal{A}$ to be the set of all distributions on Θ such that the $J-$vector $\theta_T - \theta_C$ is normally distributed. Agents in the audience must now choose a subsidy or tax for each site, so that $\mathcal{D} = \mathbb{R}^J$, and $L(d, \theta) = \|d - (\theta_T - \theta_C)\|^2$ for $\|\cdot\|$ the Euclidean norm. At each site the number of treatment and control units is equal to N, which we continue to assume is commonly known along with the variance σ^2 of weight gain. Consider two estimators, c^M and c^{JS} defined as

$$c^M(X) = \overline{X}_T - \overline{X}_C, \quad c^{JS}(X) = \max\left\{1 - \frac{J-2}{\|c^M(X)\|^2}\frac{2\sigma^2}{N}, 0\right\} c^M(X).$$

where $\overline{X}_T - \overline{X}_C$ is the $J-$vector of treatment-control differences $\overline{X}_T^j - \overline{X}_C^j$. The rule c^M corresponds to the maximum likelihood estimator for the vector of average treatment effects, while c^{JS} is a positive-part James-Stein estimator.

Claim 3. *(i) Rule c^{JS} dominates rule c^M under the decision model. (ii) Rule c^M dominates rule c^{JS} under the communication model.*

Classic results in statistics (James and Stein 1961; Baranchik 1970; Efron and Morris 1973) imply that for any value of θ the mean-squared error of rule c^{JS} is strictly lower than that of c^M, implying that c^{JS} dominates c^M under the decision model. At the same time, because $c^{JS}(X)$ is a function of $c^M(X)$, c^M is at least as good as c^{JS} for any agent under the communication model. Moreover, because c^{JS} sometimes discards useful information (by mapping a range of $\overline{X}_T - \overline{X}_C$ values to zero), c^M dominates c^{JS} under the communication model.

3.3 Generalization

The examples in this section illustrate two general properties of dominance orderings under the communication model. The first is that coarsening the analyst's report is never desirable.

Proposition 1. *Fix rules $c, c' \in \mathcal{C}$. (i) If the distribution of $c'(X)\,|c(X), X$ is equal to $\psi(c(X))$ for some $\psi : \mathcal{S} \to \Delta(\mathcal{S})$, then c achieves weakly lower risk than c' for all agents $a \in \mathcal{A}$ under the communication model. (ii) If, further, there exists $a \in \mathcal{A}$ for whom $c(X)$ and $c'(X)$ imply different optimal actions with positive probability,*

$$\Pr_a\left\{\arg\min_{d\in\mathcal{D}} \mathrm{E}_a\left[L(d,\theta)\,|c(X)\right] \cap \arg\min_{d\in\mathcal{D}} \mathrm{E}_a\left[L(d,\theta)\,|c'(X)\right] = \emptyset\right\} > 0,$$

where both minima are achieved and $R_a^*(c)$ is finite, then c dominates c' under the communication model.

The conditions in Proposition 1 part (i) imply that $c'(X)$ is a garbling of $c(X)$, while part (ii) gives a sufficient condition for strict superiority of c for a given loss function and prior. An important special case of garbling is when $c'(X)$ can be written as a deterministic transformation of $c(X)$, as in the examples in this section.

The second general property is that, following Blackwell (1951, 1953), it is optimal for the analyst to report sufficient statistics when feasible.

Proposition 2. *Fix a rule $c \in \mathcal{C}$. If $c(X)$ is sufficient for θ under F_θ, then c achieves weakly lower risk than does any other rule for all agents $a \in \mathcal{A}$ under the communication model.*

The statements about the communication model in Claims 1, 2, and 3 are corollaries of Propositions 1 and 2.

4 Conflict in Admissibility

We will say that a rule c is admissible under a given model if no other rule dominates c. Admissibility in the decision model corresponds to what Stoye (2012) terms Γ-admissibility. In this section, we give an example in which the sets of admissible rules under the decision and communication models do not intersect.

4.1 Optimal Treatment Assignment

An analyst must make a clinical recommendation to an audience of physicians on the basis of the available evidence. 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., diagnoses). 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 the analyst observes data from a trial where $n \geq 1$ 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 that patient outcomes are exchangeable, 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 θ_t denotes the success probability for units assigned to treatment t. We assume each θ_t lies in a nontrivial interval $\Theta_0 \subseteq (0, 1)$, so the parameter space is $\Theta = \Theta_0^T \subseteq (0, 1)^T$. We take the audience to consist of all possible priors $\mathcal{A} = \Delta(\Theta)$.

Each physician's decision consists of either picking a treatment t or declining to 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 physician'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).

Again consider two rules. The first rule, c^*, takes $c^*(X) = \arg\max_t X_t$ if the argmax is unique and otherwise randomizes uniformly over $\arg\max_t X_t$. The second rule, $\tilde{c}$, takes $\tilde{c}(X) = \iota$ if $\arg\max_t X_t = \{1, ..., T\}$ and $\tilde{c}(X) = c^*(X)$ otherwise. As in Section 3, the comparison of these two rules reveals a conflict in dominance ordering between the communication and decision models.

Claim 4. *(i) Rule c^* dominates rule $\tilde{c}$ under the decision model. (ii) Rule $\tilde{c}$ dominates rule c^* under the communication model.*

Start with 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 (2021). Classical decision-theoretic results for selection problems (Lehmann 1966; Eaton 1967) imply that the rule c^* minimizes decision risk uniformly over $\mathcal{A} = \Delta(\Theta)$ among rules that are invariant with respect to permutations of the treatments, and that c^* is an optimal decision rule for any agent a^* with a permutation-invariant prior.[6] By contrast, because the rule $\tilde{c}$ sometimes fails to make a recommendation, thus choosing the bad decision $d = \iota$, the rule $\tilde{c}$ is not an optimal decision rule for any agent $a \in \mathcal{A}$.

[6]The results of Stoye (2009) further imply that c^* is a minimax decision rule in the case of $T = 2$.

Next consider the communication model. Any agent can construct $c^*(X)$ given $\tilde{c}(X)$ for any $X \in \mathcal{X}$. Proposition 1 therefore implies that rule $\tilde{c}$ achieves weakly lower risk than rule c^*. Note, however, that $\tilde{c}(X)$ cannot be constructed from $c^*(X)$, because $c^*(X)$ does not inform the agent when there has been a tie. Intuitively, this results in a loss of useful information for an agent a whose prior is such that they prefer to follow the rule c^* only when the data are informative about the optimal treatment. For this reason, Proposition 1 further implies that $\tilde{c}$ dominates c^* under the communication model.

In fact, the tension between the decision and communication models is stronger than what is captured by Claim 4. Because $d = \iota$ is a bad decision, any rule that sometimes recommends it is inadmissible in decision risk. But because the decision space is too small to convey the full data, $T + 1 = |\mathcal{D}| < |\mathcal{X}| = (n+1)^T$, and distinct realizations of the data imply distinct optimal actions for some agent, any rule that does not sometimes report $c(X) = \iota$ is inadmissible in communication risk.

Claim 5. *There exists no rule c that is admissible under both the decision model and the communication model.*

We prove Claim 5 as a consequence of a more general result for situations with finite decision and sample spaces. Loosely, if there is a decision that is always unappealing and the decision space is too small to convey the actionable information in the data, then there exists no rule that is feasible and admissible in both the decision model and the communication model.

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.

4.2 Discussion

The example illustrates a tension between the decision and communication models that arises when the data are completely uninformative. Reporting that findings are inconclusive arises in many situations like the one illustrated by the UpToDate quote. Appendix B 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 priors.

Claim 5 holds for any signal space $\mathcal{S}$ containing $\mathcal{D}$. Indeed, it seems plausible that an analyst concerned with communication risk might wish to convey more than simply "I don't know." The UpToDate article that we quote at the end of Section 4.1, for example, discusses the evidence before stating its conclusion, noting that some evidence in favor of a particular selection of SSRIs failed to replicate in a second meta-analysis, and that "randomized trials have found no evidence that one antidepressant [SSRI] is superior in preventing relapse or recurrence" (Simon 2019).

The conclusion of Claim 5 also holds if we restrict $\mathcal{D}$ to contain only the feasible treatments $\{1, ..., T\}$, provided that the signal space $\mathcal{S}$ contains at least one element that is not in $\{1, ..., T\}$. Intuitively, in this case the rule $\tilde{c}$ is simply infeasible under the decision model, but remains superior to the (feasible) rule c^* under the communication model.

5 Additional Optimality Criteria

In this section we look beyond dominance comparisons to consider two other optimality criteria: optimality in weighted average risk, and minimaxity. To derive our results, we impose the following regularity conditions.

Assumption 1. *There exists a σ-finite measure which dominates F_θ for all $\theta \in \Theta$. The loss function $L(d, \theta)$ is nonnegative and lower semicontinuous in d for all $\theta \in \Theta$.*

The existence of a dominating measure is a weak condition that holds in all of our examples. Likewise, the loss functions in our examples are continuous in d, which implies lower semicontinuity.

Assumption 2. *The decision space is a subset of Euclidean space, $\mathcal{D} \subseteq \mathbb{R}^q$ for q finite, and is closed. Moreover, either (i) $\mathcal{D}$ is bounded or (ii) $\lim_{\|d\|\to\infty} L(d,\theta) = \infty$ for all θ.*

Assumption 2 holds in all of our examples. See Assumption 3 in Appendix A for a weaker condition sufficient for our results.

5.1 Weighted Average Risk

Let ω be a distribution on $\mathcal{A}$ and define

$$\rho_\omega(c) = \int_{\mathcal{A}} R_a(c)\, d\omega(a)\,,\ \ \rho^*_\omega(c) = \int_{\mathcal{A}} R^*_a(c)\, d\omega(a)$$

to be the *weighted average decision risk* and the *weighted average communication risk* of rule c, respectively. Any rule $c \in \mathcal{B}$ that minimizes weighted average decision risk $\rho_\omega(c)$ is a Bayes decision rule (e.g., Lehmann and Casella 1998, p. 6; Robert 2007, p. 63). Bayes decision rules have strong optimality properties in the classical setting.[7]

For given weights ω, weighted average risk defines a complete ordering on the set of rules, whereas dominance and admissibility define only partial orderings. These orderings are closely related.

Proposition 3. *Suppose Assumptions 1 and 2 hold. (i) If, under a given model, rule c dominates rule c', and the risk function for c is bounded and continuous in a, then c has strictly lower weighted average risk than c' with respect to any weights ω with full support on $\mathcal{A}$. (ii) If the risk functions for all rules are bounded and continuous in a, then any rule that minimizes weighted average risk with respect to full-support weights ω is admissible.*

Intuitively, if c dominates c', then at least one agent a is worse off under c than under c', and no agent is better off. As long as ω puts weight on agents in a neighborhood of a, and agents in that neighborhood have risk similar to a's, c will be strictly preferred to c' under weighted average risk.

[7] In particular, Complete Class Theorems show that in many cases any rule that cannot be expressed as Bayes is dominated by one that can be.

An implication of Proposition 3 is that if there is a conflict in dominance ordering (as in Section 3) or a conflict in admissibility (as in Section 4) between the communication and decision models, then (under the given conditions) there is a conflict in the ordering of weighted average risks with respect to full-support weights. The following corollary illustrates for the case of a conflict in admissibility.

Corollary 1. *Suppose that decision and communication risk are bounded and continuous in a for all $c \in \mathcal{B}$. If there is no rule that is admissible under both the decision and communication models, then any rule c that minimizes weighted average risk for some full-support weights ω under the decision model is inadmissible, and does not minimize weighted average risk under any full-support weights ω^*, under the communication model.*

Under the conditions of Corollary 1, any Bayes decision rule based on full-support weights ω is inadmissible for communication, and does not minimize weighted average communication risk for *any* full-support weights, including weights $\omega^* \neq \omega$.

5.2 Maximum Risk

We will say that a rule c^* is *minimax* under a given model if it minimizes the maximum risk possible under the set of priors in the audience. Formally, rule c^* is minimax if

$$R_a(c^*) = \inf_{c \in \mathcal{B}} \sup_{a \in \mathcal{A}} R_a(c), \;\; R_a^*(c^*) = \inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} R_a^*(c)$$

under the decision and communication models, respectively. Since we evaluate performance with respect to a class of priors, minimaxity in the decision model corresponds to robust Bayes optimality (also called Γ-minimaxity – see, e.g., Gilboa and Schmeidler 1989; Stoye 2012).

The max-min inequality implies that $\inf_{c \in \mathcal{B}} \sup_{a \in \mathcal{A}} R_a(c) \geq \sup_{a \in \mathcal{A}} \inf_{c \in \mathcal{B}} R_a(c)$. If the reverse is true, so that $\inf_{c \in \mathcal{B}} \sup_{a \in \mathcal{A}} R_a(c) = \sup_{a \in \mathcal{A}} \inf_{c \in \mathcal{B}} R_a(c)$, we will say that a minimax theorem holds under the decision model.[8]

[8] Viewing the decision model as a zero-sum game between the analyst and nature, a minimax theorem holds if and only if this game has a value (von Neumann and Morgenstern, 1944).

Theorem 1. *If a minimax theorem holds under the decision model, then any rule c^* that is minimax under the decision model is minimax under the communication model.*

Proof. By the definitions of decision and communication risk, for all $a \in \mathcal{A}$, $\inf_{c' \in \mathcal{B}} R_a(c') \leq R_a^*(c)$ for all $c \in \mathcal{C}$, and $R_a^*(c) \leq R_a(c)$ for all $c \in \mathcal{B}$. By the first inequality and the max-min inequality, $\sup_{a \in \mathcal{A}} \inf_{c \in \mathcal{B}} R_a(c) \leq \sup_{a \in \mathcal{A}} \inf_{c \in \mathcal{C}} R_a^*(c) \leq \inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} R_a^*(c)$. Since $R_a^*(c) \leq R_a(c)$, however, $\inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} R_a^*(c) \leq \inf_{c \in \mathcal{B}} \sup_{a \in \mathcal{A}} R_a(c)$. If a minimax theorem holds under the decision model, this implies that $\inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} R_a^*(c) = \inf_{c \in \mathcal{B}} \sup_{a \in \mathcal{A}} R_a(c)$, and any rule c^* with $R_a(c^*) = \inf_{c \in \mathcal{B}} \sup_{a \in \mathcal{A}} R_a(c)$ must also have $R_a(c^*) = \inf_{c \in \mathcal{C}} \sup_{a \in \mathcal{A}} R_a^*(c)$ and therefore be minimax under the communication model. $\square$

Thus if a minimax theorem holds under the decision model, there is no conflict between the decision and communication models when the analyst seeks to minimize maximum risk.

Theorem 1 holds for all $\mathcal{S} \supseteq \mathcal{D}$. Hence, for minimax communication, there is no gain from enlarging the signal space beyond $\mathcal{D}$, or from communicating information other than a recommended decision, provided a minimax theorem holds. The literature has derived minimax rules in a wide range of frequentist decision problems, and Theorem 1 implies that these will also be minimax communication rules for the maximal audience $\mathcal{A} = \Delta(\Theta)$, provided a minimax theorem holds. Theorem 1 also implies that robust Bayes decision rules with respect to a class of priors $\mathcal{A}$ are robust communication rules with respect to the same class of priors, provided a minimax theorem holds.

The next proposition, proved in Appendix A as a consequence of a more general result building on arguments from Strasser (1985), gives sufficient conditions for a minimax theorem to hold. To state the proposition, we say that $\mathcal{A}$ is convex if for any $a, a' \in \mathcal{A}$ and any $\lambda \in (0,1)$, $\lambda \cdot a + (1-\lambda) \cdot a' \in \mathcal{A}$.

Proposition 4. *Suppose Assumptions 1 and 2 hold. If the audience $\mathcal{A}$ is convex, then a minimax theorem holds under the decision model, and there exists a minimax rule c^*.*

Convexity of $\mathcal{A}$ holds for the maximal audience $\mathcal{A} = \Delta(\Theta)$, as well as for the classes of priors studied in Gilboa and Schmeidler (1989). We next discuss an example that illustrates the differences between minimizing weighted average risk and minimizing maximum risk, and also highlights the role played by convexity of $\mathcal{A}$ in Proposition 4.

5.3 Combining Multiple Moments

An analyst observes data $X \in \mathbb{R}^k$ and is interested in a scalar parameter τ. An economic model imposed by the analyst implies a sample moment function $g(\tilde{\tau}) = X - G\tilde{\tau}$, for G a known, nonrandom $k-$vector. Under the analyst's model the sample moments $g(\tau)$ have mean 0 at the true value of τ. The economic model may be misspecified, however, so that the true mean of $g(\tau)$ is given by (η', γ'), where (η, γ) are nuisance parameters with $\dim(\eta) + \dim(\gamma) = k$. Thus $\theta = (\tau, \eta, \gamma)$. We formalize the idea that τ is the parameter of interest by taking $\mathcal{D} = \mathbb{R}$ and $L(d, \theta) = (d - \tau)^2$.

We further assume that

$$g(\tau) \sim N\left((\eta', \gamma'), \sigma^2 \cdot I_k\right) \tag{1}$$

where $\sigma^2 > 0$ is a commonly known variance and I_k is the identity matrix. Armstrong and Kolesár (2021) show that (under regularity conditions) locally misspecified moment condition models, including nonlinear models with multiple parameters, are asymptotically equivalent to a version of (1) where the moment function depends on multiple parameters and has an unrestricted, but known, variance matrix.[9]

Suppose all agents $a \in \mathcal{A}$ believe that θ follows a multivariate normal distribution with $\tau \perp (\eta', \gamma')$ and $\tau \sim N(0, \rho_\tau^2)$ for $\rho_\tau > 0$. Agents are concerned about misspecification, and each agent a believes that $\eta = \eta_a$ and $\gamma \sim N\left(0, \rho_\gamma^2 \cdot I_{\dim(\gamma)}\right)$ for $\rho_\gamma \geq 0$. Thus, agents are certain, but may disagree, about the extent of misspecification of the first $\dim(\eta)$ moments, and are uncertain, in a commonly-agreed way, about the extent of misspecification of the remaining $\dim(\gamma)$ moments.

[9]Starting with the representation in Armstrong and Kolesár (2021, equation 4), one can obtain (1) by partialling out model parameters other than τ, and then normalizing the variance of the sample moments.

We will consider an audience $\mathcal{A}$ and weights ω such that beliefs about η follow $\eta_a \sim N\left(0, \rho_\eta^2 \cdot I_{\dim(\eta)}\right)$ under ω, for $\rho_\eta > 0$. Under such weights we can characterize rules that minimize weighted average risk under both the decision and communication models. Towards such a characterization, decompose $X = \left(X_\eta', X_\gamma'\right)$ and $G = \left(G_\eta', G_\gamma'\right)$ conformably with (η', γ') and assume that $G_\eta, G_\gamma \neq 0$. Let $\hat{\tau}_\eta = \left(G_\eta' G_\eta\right)^{-1} G_\eta' X_\eta$ be the maximum likelihood estimate (absent misspecification) based on the first $\dim(\eta)$ moments, and define $\hat{\tau}_\gamma$ analogously. Note that $\operatorname{Var}(\hat{\tau}_\eta|\theta) = \sigma_\eta^2 = \sigma^2 \cdot \psi_\eta$ for $\psi_\eta = \left(G_\eta' G_\eta\right)^{-1}$, and likewise for $\hat{\tau}_\gamma$.

Claim 6. *Any rule c_ω that minimizes weighted average risk under the decision model takes*

$$c_\omega(X) = c_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma) = \frac{\left(\sigma_\eta^2 + \rho_\eta^2 \psi_\eta\right)^{-1} \cdot \hat{\tau}_\eta + \left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1} \cdot \hat{\tau}_\gamma}{\sigma_\tau^{-2} + \left(\sigma_\eta^2 + \rho_\eta^2 \psi_\eta\right)^{-1} + \left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1}}$$

almost surely. One rule c_ω^* that minimizes weighted average risk under the communication model takes

$$c_\omega^*(X) = c_\omega^*(\hat{\tau}_\eta, \hat{\tau}_\gamma) = \frac{\sigma_\eta^{-2} \cdot \hat{\tau}_\eta + \left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1} \cdot \hat{\tau}_\gamma}{\sigma_\tau^{-2} + \sigma_\eta^{-2} + \left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1}}.$$

The rule c_ω does not minimize weighted average risk under the communication model, and the rule c_ω^* does not minimize weighted average risk under the decision model. Moreover, no other rule achieves strictly lower communication risk than c_ω^* for any agent.

Both $c_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ and $c_\omega^*(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ are weighted averages of the prior mean (i.e., 0) and the maximum likelihood estimates $(\hat{\tau}_\eta, \hat{\tau}_\gamma)$. The weighted averages differ in the weight they place on $\hat{\tau}_\eta$, the maximum likelihood estimate based on the block of moments about which the agents disagree. Each agent a is confident that the bias of $\hat{\tau}_\eta$ for τ is $\left(G_\eta' G_\eta\right)^{-1} G_\eta' \eta_a$. Under the decision model, disagreement about the magnitude of the bias translates into a larger expected distance between $\hat{\tau}_\eta$ and τ. Under the communication model, such disagreement is irrelevant, because each agent a can readily compute their posterior mean

$$\mathrm{E}_a\left[\tau | c_\omega^*(\hat{\tau}_\eta, \hat{\tau}_\gamma)\right] = c_\omega^*(\hat{\tau}_\eta, \hat{\tau}_\gamma) - \frac{\sigma_\eta^{-2} \left(G_\eta' G_\eta\right)^{-1} G_\eta' \eta_a}{\sigma_\tau^{-2} + \sigma_\eta^{-2} + \left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1}}$$

that adjusts $c^*_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ for the bias in $\hat{\tau}_\eta$. As a result, $c^*_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ places more weight on $\hat{\tau}_\eta$ than does $c_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$. And because $\mathrm{E}_a[\tau | c^*_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)]$ coincides with agent a's posterior mean $\mathrm{E}_a[\tau|X]$ based on the full data, no communication rule can be better than c^*_ω from the agent's point of view.

In the limit taking disagreement to zero, $\rho_\eta \to 0$, the two rules coincide, whereas in the limit taking disagreement to infinity, $\rho_\eta \to \infty$, $c_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ places no weight on $\hat{\tau}_\eta$, and $c^*_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ is unaffected. By contrast, the rules c_ω and c^*_ω are similar in how they treat agents' uncertainty about γ, and in the limit as $\rho_\gamma \to \infty$ both rules place no weight on $\hat{\tau}_\gamma$. Thus, the decision and communication models both predict that the analyst will down-weight moments about whose validity the audience is very uncertain. In contrast to the communication model, however, the decision model further predicts that the analyst will down-weight moments about whose misspecification audience members disagree, even if each audience member is certain in their belief.

In practice analysts frequently choose from among a large set of potentially misspecified moments when estimating economic models. Nakamura and Steinsson (2018) advocate estimating structural models of the macroeconomy by targeting "identified moments" that correspond to direct estimates of causal effects (see also Dridi et al. 2007). Nakamura and Steinsson (2018) argue that, although the assumptions justifying the causal interpretation of identified moments "are typically controversial," these moments are sensitive to a relatively narrow range of modeling assumptions. By contrast, other moments one could target, for example unconditional means and variances, are likely to be sensitive to the specification of many different aspects of the model. Targeting identified moments may therefore allow audience members to form more precise beliefs about the likely impact of misspecification on the analyst's estimate. In this sense the recommendation to target identified moments, in preference to moments whose behavior under misspecification is harder to assess, seems more consistent with the predictions of the communication model than with those of the decision model.

It is also possible to characterize minimax rules in this example.

Claim 7. *Any rule $\check{c}_\omega$ that is minimax under the decision model takes*

$$\check{c}_\omega(X) = \check{c}_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma) = \frac{\left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1} \cdot \hat{\tau}_\gamma}{\sigma_\tau^{-2} + \left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1}}$$

almost surely, and therefore coincides with c_ω in the limit as $\rho_\eta \to \infty$. The rule $\check{c}_\omega$ is not minimax under the communication model.

Because any decision rule that puts weight on $\hat{\tau}_\eta$ can be arbitrarily bad for sufficiently large η_a, the rule $\check{c}_\omega$ puts no weight on $\hat{\tau}_\eta$. Ignoring $\hat{\tau}_\eta$ is unappealing under the communication model, however, because $\hat{\tau}_\eta$ is informative about τ, and as discussed above agents can account for the bias in $\hat{\tau}_\eta$ in formulating their optimal decision, no matter how large is η_a.

Proposition 4 does not apply in this setting because the audience $\mathcal{A}$ is not convex. If we replace the audience $\mathcal{A}$ with its convex hull, then by Proposition 4 and Theorem 1, $\check{c}_\omega$ is a minimax rule under the communication model. Intuitively, convexifying the audience means that if there exist agents a and a' who disagree about η, there exists a third agent a'' who puts equal weight on the two beliefs. Convexity thus turns disagreement about how to interpret $\hat{\tau}_\eta$ into uncertainty, and so implies that minimax communication rules should put no weight on $\hat{\tau}_\eta$. This illustrates the role of the convexity restriction on $\mathcal{A}$ in Proposition 4.

Discussion

Under the communication model, rule c_ω^* is more appealing than rule c_ω because agents can adjust for the bias in $\hat{\tau}_\eta$ when forming their own decisions or judgments. To make the appropriate adjustment, agents need to know the weight that $c_\omega^*(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ places on $\hat{\tau}$. Andrews et al. (2020) discuss the situation where weights are data-dependent, in which case it is appealing (from the standpoint of communication risk) for the analyst to report the weights to the audience.

Under the communication model, the rule $c_\omega^*(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ is as good, for any agent, as having access to the full data X.[10] This property of the example depends on the

[10] In particular, $c_\omega^*(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ is marginally sufficient for τ with respect to $\mathcal{A}$ in the sense of Raiffa and Schlaifer (1961).

assumption that all agents have the same prior variance for θ. In situations with more heterogeneity in agents' beliefs, there need not be a low-dimensional sufficient statistic. Appendix C considers a setting where each component of $g(\tau)$ is subject to an additional disturbance, on which agents have mean-zero Gaussian priors with potentially different prior variances. In this case, any coarsening of the data increases communication risk for some agent, and an analyst concerned with communication risk in such a setting might be expected to report X to the audience. DellaVigna (2018) advocates reporting X as good practice when estimating structural models in behavioral economics. Appendix C shows that, as the size of the additional disturbance becomes small, $c^*_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ achieves communication risk arbitrarily close to that from observing X. If there are communication constraints (say because k is large or some data must remain confidential), $c^*_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ therefore remains appealing under the communication model.

6 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 exhibit settings in which the proposed model predicts very different reporting rules from the classical model. We argue that, in some practical situations similar to these settings, scientists' reports appear more consistent with the predictions of the proposed model than with the predictions of the classical model.

A Proofs

Proofs of Claims

Proof of Claim 1. To see that c' dominates c under the decision model, note that $\Pr_a\{c(X)<0\}>0$ for all $a\in\mathcal{A}$, and

$$\mathrm{E}_a\left[L\left(c\left(X\right),\theta\right)|c\left(X\right)<0\right]>\mathrm{E}_a\left[L\left(c^{'}\left(X\right),\theta\right)|c\left(X\right)<0\right],$$

while the two rules achieve the same loss when $c(X)\geq 0$. Dominance in the decision model follows immediately.

For dominance in the communication model consider an agent with a $N(\mu,1)$ prior on $\theta_T-\theta_C$, truncated at zero. This agent's posterior on $\theta_T-\theta_C$ after observing $c(X)=d$ is a $N\left(\frac{1}{1+\tilde{\sigma}^{-2}}\mu+\frac{\tilde{\sigma}^{-2}}{1+\tilde{\sigma}^{-2}}d,\left(1+\tilde{\sigma}^{-2}\right)^{-1}\right)$ truncated at zero, for $\tilde{\sigma}^2=\frac{\sigma^2}{N_T}+\frac{\sigma^2}{N_C}$. This agent's optimal decision is thus a strictly increasing function of $c(X)$. Since $c'(X)$ is a non-invertible transformation of $c(X)$, and $\arg\min_{d\in\mathcal{D}}\mathrm{E}_a\left[L(d,\theta)|c'(X)\right]$ is a singleton by strict convexity of the loss, for almost every $\tilde{d}<0$

$$\arg\min_{d\in\mathcal{D}}\mathrm{E}_a\left[L\left(d,\theta\right)|c\left(X\right)=\tilde{d}\right]\cap\arg\min_{d\in\mathcal{D}}\mathrm{E}_a\left[L\left(d,\theta\right)|c'\left(X\right)=0\right]=\emptyset.$$

Proposition 1 thus implies that c dominates c' in the communication model. $\square$

Proof of Claim 2. For part (i) of the claim, consider an agent with a dogmatic prior that $\theta_C=0$ with probability one. Suppose further that this agent has a $N(\mu,1)$ prior on θ_T. This agent's posterior on the average treatment effect after observing $\left(\bar{X}_C,\bar{X}_T\right)$ will be a $N\left(\frac{1}{1+\tilde{\sigma}_T^{-2}}\mu+\frac{\tilde{\sigma}_T^{-2}}{1+\tilde{\sigma}_T^{-2}}\bar{X}_T,\left(1+\tilde{\sigma}_T^{-2}\right)^{-1}\right)$ distribution truncated at zero, for $\tilde{\sigma}_T^2=\frac{\sigma^2}{N_T}$. Hence, this agent's optimal action is a strictly increasing transformation of $\bar{X}_T$. Under the agent's prior, however, $c(X)$ is equal to $\bar{X}_T$ plus standard normal noise, so the agent cannot implement this optimal action based on observing $c(X)$ alone. Proposition 1 thus implies that c'' dominates c under the communication model.

Part (ii) of the claim is immediate from Proposition 2. $\square$

Proof of Claim 3. Part (i) of the claim follows from standard results on the positive-part James-Stein estimator (see e.g. Efron and Morris 1973). For part (ii) of the claim, consider an agent with a $N(\mu, I_J)$ prior on the vector of average treatment effects, and note that this agent's posterior mean is a one-to-one transformation of $c(X) = \bar{X}_T - \bar{X}_C$. Hence, the conclusion follows from Proposition 1 by an argument similar to the proof of Claim 1. $\square$

Proof of Claim 4. Under the decision model, all agents $a \in \mathcal{A}$ strictly prefer to randomize uniformly over $d \in \{1, ..., T\}$ rather than taking $d = \iota$. That c^* dominates $\tilde{c}$ in the decision model follows immediately.

For part (ii) of the claim, note that $\tilde{c}$ yields weakly smaller communication risk for all agents than c^* by part (i) of Proposition 1. To show strict inequality for some agents, consider agents a for whom $\mathrm{E}_a [\theta_t | \arg\max_t X_t = \{1, ..., T\}]$ is non-constant across t, while for all X $\arg\max_t \mathrm{E}_a [\theta_t | c^*(X)] = c^*(X)$. Any agent a^* with a permutation-invariant prior has $c^*(X) \in \arg\max_t \mathrm{E}_{a^*} [\theta_t | c^*(X)]$, so we can find agents a of the sort we desire by slightly perturbing such a prior. 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. Dominance in the communication model follows by Proposition 1. $\square$

Definition 1. *Suppose $\mathcal{X}$ is finite, and 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,*

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

The ***effective size of the sample space*** *$\mathcal{X}$, denoted $N(\mathcal{X}, \mathcal{A})$, is the minimal size of a partition in $\mathcal{P}^*$, $N(\mathcal{X}, \mathcal{A}) = \min\{|P| : P \in \mathcal{P}^*\}$.*

Proposition 5. *Suppose that $\mathcal{D}$ and $\mathcal{X}$ are finite, that $L(d, \theta)$ is bounded, and that there exists a decision $d \in \mathcal{D}$ with $L(d, \theta) \geq L(d', \theta)$ for all $\theta \in \Theta$ and some $d' \in \mathcal{D}$, with strict inequality for all $\theta \in \tilde{\Theta} \subseteq \Theta$. Suppose further that $\Pr_a \left\{ \tilde{\Theta} \right\} > 0$ for some $a \in \mathcal{A}$, and that F_θ has support $\mathcal{X}$ for all $\theta \in \Theta$. If $N(\mathcal{X}, \mathcal{A}) \geq |\mathcal{D}|$, then any rule c that is admissible in decision risk is inadmissible in communication risk and vice versa.*

Proof of Claim 5. We prove this result building on Proposition 5. First, note that choosing $d' = 1$ yields strictly lower loss than choosing $d = \iota$ for all $\theta \in \Theta$, which verifies the condition on the loss. Next, note that the effective size $N(\mathcal{X}, \Delta(\Theta))$ of the sample space is bounded below by the size $N\left(\mathcal{X}, \tilde{\mathcal{A}}\right)$ for a restricted audience $\tilde{\mathcal{A}} \subseteq \Delta(\Theta)$. Consider the audience consisting of only three agents, a_0, a_1, and a_2. Agent a_0 has a uniform prior on Θ. This implies that θ_t is independent of θ_s for all $s \neq t$. By the monotone likelihood ratio property of the binomial distribution, 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 not 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

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

for $f(X;\theta)$ the probability mass function of F_θ, where F_θ has full support for all $\theta \in \Theta$. Hence, $\mathrm{E}_a[\theta|X]$ is continuous in a (for the L_1 norm on $\mathcal{A}$). Thus, 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 strictly prefers $d = 1$ when $\arg\max_t X_t = \{1, ..., T\}$, and an agent a_2 who strictly prefers $d = 2$ conditional on the same event. This immediately implies, however, that $N\left(\mathcal{X}, \tilde{\mathcal{A}}\right) \geq T + 1$, 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. □

Proof of Claim 6. Note that $\rho_\omega(c) = R_{a_\omega}(c)$ for $a_\omega\left(\tilde{\Theta}\right) = \int_{\mathcal{A}} a\left(\tilde{\Theta}\right) d\omega(a)$ for all $\tilde{\Theta} \subseteq \Theta$. Hence, weighted average decision risk is simply the decision risk for the agent with the weighted average prior, a_ω. However, the rule $c_\omega(X)$ corresponds to the posterior mean for this prior, and hence is the almost-surely unique optimal rule in the decision model.

For the communication model, by contrast, note that $\mathrm{E}_a\left[\tau | c_\omega^*\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)\right] = \mathrm{E}_a\left[\tau | X\right]$ corresponds to agent a's posterior mean. Hence, $c_\omega^*\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)$ allows all agents to obtain the same risk as if they observed the full data, and so is optimal in the communication model. By contrast, for all agents $a \in \mathcal{A}$, $\mathrm{Var}_a\left(c_\omega^*\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right) | c_\omega\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)\right) > 0$, so $c_\omega^*\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)$ has strictly lower weighted average risk than $c_\omega\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)$ under the communication model. $\square$

Proof of Claim 7. Note that for all c, $\sup_{a \in \mathcal{A}} R_a(c) = \sup_\omega \rho_\omega(c)$. Hence, to obtain a minimax decision rule, we want to solve $\min_{c \in \mathcal{B}} \sup_\omega \rho_\omega(c)$. Note, next, that the decision risk of $\check{c}_\omega\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)$ is the same for all $a \in \mathcal{A}$, and that $\check{c}_\omega\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)$ corresponds to a Bayes decision rule for an agent with an infinite-variance normal prior on η, and independent $N\left(0, \sigma_\tau^2\right)$ and $N\left(0, \rho_\gamma^2 \cdot I_{\dim(\gamma)}\right)$ priors on τ and γ, respectively. Denote the corresponding (limit of) weights by ω^*, and note that for all $a \in \mathcal{A}$,

$$R_a\left(\check{c}_\omega\right) = \rho_{\omega^*}\left(\check{c}_\omega\right) = \min_{c \in \mathcal{B}} \rho_{\omega^*}(c).$$

Since $\rho_\omega\left(\check{c}_\omega\right) = \rho_{\omega^*}\left(\check{c}_\omega\right)$ for all ω, it follows immediately that $\check{c}_\omega$ is a minimax decision rule. Since the loss function is strictly convex, it is almost surely unique. Finally, building on the proof of Claim 6, note that $\mathrm{E}_a\left[\tau | \check{c}_\omega\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)\right] = \frac{\left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1} \cdot \hat{\tau}_\gamma}{\sigma_\tau^{-2} + \left(\sigma_\gamma^2 + \rho_\gamma^2 \psi_\gamma\right)^{-1}}$ for all a, and that $\mathrm{E}_a\left[\left(\mathrm{E}_a\left[\tau | \check{c}_\omega\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)\right] - \mathrm{E}_a\left[\tau | c_\omega^*\left(\hat{\tau}_\eta, \hat{\tau}_\gamma\right)\right]\right)^2\right] = \varepsilon > 0$ for a constant ε. Hence $\check{c}_\omega$ is not minimax under the communication model. $\square$

Proofs of Propositions

Proof of Proposition 1. For part (i) of the proposition, under the garbling condition, an agent who observes $c(X)$ can generate draws from the distribution of $c'(X) | c(X), X$ by applying ψ to the observed report $c(X)$. This, however, implies that $\mathrm{E}_a\left[L(d, \theta) | c(X), c'(X)\right] = \mathrm{E}_a\left[L(d, \theta) | c(X)\right]$, so

$$R_a(c) = \mathrm{E}_a\left[\inf_{d \in \mathcal{D}} \mathrm{E}_a\left[L(d, \theta) | c(X)\right]\right] \leq \mathrm{E}_a\left[\inf_{d \in \mathcal{D}} \mathrm{E}_a\left[L(d, \theta) | c'(X)\right]\right] = R_a(c').$$

For part (ii) of the proposition, let us write $\mathcal{E} \subseteq \mathcal{X}$ for the event that

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

Note that $\mathrm{E}_a\left[L(d,\theta)\,|c(X),\mathcal{E}\right] = \mathrm{E}_a\left[L(d,\theta)\,|c(X)\right]$, and consider $f : \mathcal{X} \to \mathcal{D}$ such that $f(X)$ lies in $\arg\min_{d\in\mathcal{D}} \mathrm{E}_a\left[L(d,\theta)\,|c'(X)\right]$ almost surely. By definition, $\mathrm{E}_a\left[L(f(X),\theta) - \min_{d\in\mathcal{D}} \mathrm{E}_a\left[L(d,\theta)\,|c(X)\right]|\mathcal{E}\right] > 0$, so since

$$\mathrm{E}_a\left[L(f(X),\theta) - \min_{d\in\mathcal{D}} \mathrm{E}_a\left[L(d,\theta)\,|c(X)\right]|\mathcal{X}\setminus\mathcal{E}\right] \geq 0,$$

the result follows. □

Proof of Proposition 2. Sufficiency of $c(X)$ implies that for any other report $c'(X)$ and any prior a, the distribution of $\theta|c(X), c'(X)$ is the same as that of $\theta|c(X)$. Hence, $\mathrm{E}_a\left[L(d,\theta)\,|c(X), c'(X)\right] = \mathrm{E}_a\left[L(d,\theta)\,|c(X)\right]$, and the argument is the same as in part (i) of Proposition 1. □

Assumption 3. *Either (i) $\mathcal{D}$ is compact or (ii) $\mathcal{D}$ is locally compact with a countable base, and $\{d : L(d,\theta) \leq l\}$ is compact for all $l \in \mathbb{R}$ and $\theta \in \Theta$.*

Lemma 1. *Under Assumption 1, Assumption 2 implies Assumption 3.*

Proof of Lemma 1. Case (i) of Assumption 2 trivially implies case (i) of Assumption 3. For case (ii), closed subsets of Euclidean spaces are locally compact, and lower semi-continuity of L implies that $\{d : L(d,\theta) \leq l\}$ is closed for all l. Assumption 2 (ii) implies that $\{d : L(d,\theta) \leq l\}$ is bounded. □

Lemmas 2 and 3, Proposition 4, Corollary 2, or their proofs, consider generalized decision functions. For $\mathcal{H}$ the space of bounded continuous functions $h : \mathcal{D} \to \mathbb{R}$, and $\mathcal{M}$ the set of bounded signed measures on $\mathcal{X}$, define the class of generalized decision functions $\mathcal{G}$ as the set of bilinear functions $g : \mathcal{H} \times \mathcal{M} \to \mathbb{R}$ with (i) $|g(h,\mu)| \leq \|h\|_\infty \|\mu\|_1$, (ii) $g(h,\mu) \geq 0$ if $h \geq 0$ and $\mu \geq 0$, and (iii) $g(1,\mu) = \|\mu\|_1$ if $\mu \geq 0$. For $c \in \mathcal{B}$, let $c(\cdot;x)$ be the measure on $\mathcal{D}$ implied by $c(x)$, define $g_c(h,\mu) = \int_\mathcal{X}\int_\mathcal{D} h(d)dc(d;x)\,d\mu(x)$, and note that $\{g_c : c \in \mathcal{B}\} \subseteq \mathcal{G}$. Further, for $\ell_\theta : \mathcal{D} \to \mathbb{R}$ define $W(g,\ell,a) = \int_\Theta g(\ell_\theta, F_\theta)\,da(\theta)$, and note that $W(g_c, L, a) = R_a(c)$.

Lemma 2. *Under Assumption 1, decision risk is lower semicontinuous in a for all $c \in \mathcal{C}$. Under Assumptions 1 and 3, the same holds for communication risk.*

Proof of Lemma 2. Theorem 42.3 of Strasser (1985) establishes that $\mathcal{G}$ is convex, and compact in the weak topology (i.e. the topology such that $g_k \to g$ if and only if $g_k(h,\mu) \to g(h,\mu)$ for all $(h,\mu) \in \mathcal{H} \times \mathcal{M}$). Let $L_\theta(d) = L(d,\theta)$. Since $L(d,\theta)$ is lower semicontinuous in d for all θ, Lemma 47.2 of Strasser (1985) establishes that $g(L_\theta,\mu) = \sup_{\ell \in \mathcal{L}_\theta} g(\ell,\mu)$ for $\mathcal{L}_\theta$ the set of bounded, nonnegative, and continuous functions $\ell : \mathcal{D} \to \mathbb{R}$ with $\ell \leq L_\theta$. Note that $g(\ell,\mu)$ is continuous with respect to the product of the weak topology on $\mathcal{G}$ and the L_1 topology on $\mathcal{M}$.

Next, define $\tilde{\mathcal{L}}$ to be the set of functions $\tilde{\ell}$ with $\tilde{\ell}_\theta \in \mathcal{L}_\theta$ for all θ and $\sup_{\theta,d} \tilde{\ell}_\theta(d)$ finite, and let $W\left(g,\tilde{\ell},a\right) = \int_\Theta g\left(\tilde{\ell}_\theta, F_\theta\right) da(\theta)$. $W\left(g,\tilde{\ell},a\right)$ is lower semi-continuous with respect to the product of the weak topology on $\mathcal{G}$ and the L_1 topology on $\mathcal{A}$. Since $\mathcal{G}$ is compact in the weak topology, the theorem of the maximum implies that $\inf_{g\in\mathcal{G}} W\left(g,\tilde{\ell},a\right)$ is likewise lower semicontinuous on $\mathcal{A}$ (see Lemma 17.30 in Aliprantis and Border, 2006).

The supremum of a family of lower semicontinuous functions remains lower semicontinuous, so both $\sup_{\tilde{\ell}\in\tilde{\mathcal{L}}} W\left(g,\tilde{\ell},a\right)$ and $\sup_{\tilde{\ell}\in\tilde{\mathcal{L}}} \inf_{g\in\mathcal{G}} W\left(g,\tilde{\ell},a\right)$ are lower semicontinuous in a. For the former, note that $\sup_{\tilde{\ell}\in\tilde{\mathcal{L}}} W\left(g,\tilde{\ell},a\right) = W(g,L,a)$ (again by Lemma 47.2 of Strasser, 1985). For the latter, note that $\tilde{\mathcal{L}}$ is convex, while $\mathcal{G}$ is convex and compact in the weak topology. $W\left(g,\tilde{\ell},a\right)$ is lower semicontinuous in g and continuous in $\tilde{\ell}$ (for the uniform topology on $\tilde{\mathcal{L}}$). Hence, Sion's (1958) minimax theorem (Corollary 3.3 in Sion 1958) implies that

$$\sup_{\tilde{\ell}\in\tilde{\mathcal{L}}} \inf_{g\in\mathcal{G}} W\left(g,\tilde{\ell},a\right) = \inf_{g\in\mathcal{G}} \sup_{\tilde{\ell}\in\tilde{\mathcal{L}}} W\left(g,\tilde{\ell},a\right) = \inf_{g\in\mathcal{G}} W(g,L,a).$$

Hence, $\inf_{g\in\mathcal{G}} W(g,L,a)$ is lower semicontinuous in a.

To complete the proof, we need to relate these results back to attainable risk functions. For decision risk, recall that $\{g_c : c \in \mathcal{B}\} \subseteq \mathcal{G}$ and note that $W(g_c,L,a) = R_a(c)$, so we have proved lower semicontinuity of R_a in a.

For communication risk, consider case (i) in Assumption 3. Theorem 43.2 of Strasser (1985) implies that for each $g \in \mathcal{G}$, there exists some $c \in \mathcal{B}$ with $g_c(h,\mu) =$

$g(h, \mu)$ for all $\mu \geq 0$ and all lower semicontinuous functions h that are bounded from below. Hence, $\inf_{g \in \mathcal{G}} W(g, L, a) = \inf_{c \in \mathcal{C}} W(g_c, L, a)$. Next consider case (ii). Theorem 43.5 of Strasser (1985) implies that for each $g \in \mathcal{G}$, there exists $c \in \mathcal{B}$ such that $g_c(h, \mu) \leq g(h, \mu)$ for all h with compact sublevel sets $\{d : h(d) \leq l\}$. Hence, again $\inf_{c \in \mathcal{C}} W(g_c, L, a) = \inf_{g \in \mathcal{G}} W(g, L, a)$. However, $\inf_{c \in \mathcal{C}} W(g_c, L, a)$ is equal to the communication risk based on observing the full data, so the conclusion is immediate by considering the special case where the data X are reduced to just the analyst's report. $\square$

Proof of Proposition 3. We discuss the argument for the decision model, while the result for the communication model follows by the same argument. Lemma 2 implies that $R_a(c')$ is lower semicontinuous in a, while $R_a(c)$ is continuous in a by assumption. Hence, $R_a(c') - R_a(c)$ is lower semicontinuous. Dominance of c means that $\{a : R_a(c') - R_a(c) < 0\}$ is empty, while $\{a : R_a(c') - R_a(c) > 0\}$ is nonempty, and is open by lower semicontinuity. Since ω has full support, this implies that

$$\rho_\omega(c') - \rho_\omega(c) = \int_{\mathcal{A}} 1\{R_a(c') - R_a(c) > 0\}(R_a(c') - R_a(c))\, d\omega(a) > 0.$$

Since $R_a(c)$ is bounded $\rho_\omega(c)$ is finite, proving part (i) of the proposition.

For part (ii) of the proposition, suppose towards contradiction that the rule c minimizes weighted average risk, but is dominated by another rule c''. The proof of part (i) implies that $\rho_\omega(c) > \rho_\omega(c'')$, which contradicts weighted average optimality of c. $\square$

Proof of Corollary 1. By Proposition 3, under the conditions of the corollary any rule that minimizes weighted average risk with respect to full-support weights in a given model is admissible in that model. Hence, if the set of admissible rules for the decision and communication model do not overlap, weighted average risk optimality in the decision model implies inadmissibility, and hence non-optimality in weighted average risk for any full-support weights, in the communication model. $\square$

Lemma 3. *(Extension of Lemma 46.1 in Strasser 1985) Suppose that L is bounded and continuous. For every continuous $f : \mathcal{A} \to \mathbb{R}$ the following two statements are* *equivalent: (i) there exists $g \in \mathcal{G}$ such that $f(a) \geq W(g, L, a)$ for all $a \in \mathcal{A}$, (ii) $\int f(a)\, d\omega(a) \geq \inf \left\{ \int W(g, L, a)\, d\omega(a) : g \in \mathcal{G} \right\}$ for every weight function ω on $\mathcal{A}$.*

Proof of Lemma 3. That (i) implies (ii) is immediate. To show that (ii) implies (i), note that Theorem 45.6 of Strasser (1985) (taking $M_1 = \{f\}$ and $M_2 = \{W(g, L, a) : g \in \mathcal{G}\}$) implies that for $\mathbb{C}(\mathcal{A})$ the set of continuous functions on $\mathcal{A}$, there exists some $\tilde{g}$ in the closure of $\mathcal{W} = \cup_{g \in \mathcal{G}} \{\bar{g} \in \mathbb{C}(\mathcal{A}) : \bar{g} \geq W(g, L, \cdot)\}$ with $\tilde{g} \leq f$. Theorem 42.3 of Strasser (1985) establishes that $\mathcal{G}$ is convex, and compact in the weak topology. Hence, $\mathcal{W}$ is closed by Remark 45.4 of Strasser (1985), and $\tilde{g} \in \mathcal{W}$. Thus, (ii) implies (i), and we have established equivalence. $\square$

Corollary 2. *(Extension of Corollary 46.2 in Strasser 1985) The conclusion of Lemma 3 holds for any loss function L that is lower semicontinuous in d.*

Proof of Corollary 2. That (i) implies (ii) is again immediate. To obtain (i) from (ii), define $\tilde{\mathcal{L}}$ as in the proof of Lemma 2. Condition (ii) implies that $\int f(a)\, d\omega(a) \geq \inf \left\{ \int W\left(g, \tilde{\ell}, a\right) d\omega(a) : g \in \mathcal{G} \right\}$ for all ω and all $\tilde{\ell} \in \tilde{\mathcal{L}}$. Hence, by Lemma 3, for each $\tilde{\ell} \in \tilde{\mathcal{L}}$ the set $\left\{ g \in \mathcal{G} : W\left(g, \tilde{\ell}, a\right) \leq f(a) \text{ for all } a \in \mathcal{A} \right\}$ is nonempty. Note that this set is decreasing as $\tilde{\ell}$ increases pointwise, so since L is the pointwise upper bound of $\tilde{\mathcal{L}}$, Cantor's intersection theorem implies that $\{g \in \mathcal{G} : W(g, L, a) \leq f(a) \text{ for all } a \in \mathcal{A}\}$ is nonempty, which in turn implies (i). $\square$

Proof of Proposition 4. As discussed in Section 5.2, we need only show that

$$\inf_{c \in \mathcal{B}} \sup_{a \in \mathcal{A}} R_a(c) \leq \sup_{a \in \mathcal{A}} \inf_{c \in \mathcal{B}} R_a(c).$$

To do so, note that $\sup_{a \in \mathcal{A}} \inf_{c \in \mathcal{B}} R_a(c) \leq \sup_{\omega} \inf_{c \in \mathcal{B}} \rho_\omega(c)$, and let $f(a)$ be the constant function equal to $\sup_\omega \inf_{c \in \mathcal{B}} \rho_\omega(c)$ for all a. By construction $\int f(a)\, d\omega(a) \geq \inf \left\{ \int W(g, L, a)\, d\omega(a) : g \in \mathcal{G} \right\}$ for all ω, so by Corollary 2 there exists $g^* \in \mathcal{G}$ with $f(a) \geq W(g^*, L, a)$ for all $a \in \mathcal{A}$.

For case (i) in Assumption 3, Theorem 43.2 of Strasser (1985) implies that there exists some $c^* \in \mathcal{B}$ with $g_{c^*}(L, \mu) = g^*(L, \mu)$ for all $\mu \geq 0$. For case (ii), Theorem 43.5 of Strasser (1985) implies that there exists $c^* \in \mathcal{B}$ such that $g_{c^*}(L, \mu) \leq g^*(L, \mu)$.

For these c^*, $\sup_{a\in\mathcal{A}} R_a(c^*) \leq \sup_\omega \inf_{c\in\mathcal{B}} \rho_\omega(c)$ by construction. Since $\mathcal{A}$ is convex, however, $\sup_\omega \inf_{c\in\mathcal{B}} \rho_\omega(c)$ is equal to $\sup_{a\in\mathcal{A}} \inf_{c\in\mathcal{B}} R_a(c)$, so $\inf_{c\in\mathcal{B}} \sup_{a\in\mathcal{A}} R_a(c) \leq \sup_{a\in\mathcal{A}} R_a(c^*) \leq \sup_{a\in\mathcal{A}} \inf_{c\in\mathcal{B}} R_a(c)$ and c^* is a minimax rule under the decision model. $\square$

Proof of Proposition 5. We first argue that any rule c that is admissible in decision risk must use the decision d with probability zero. Specifically, consider any a with $\Pr_a\left\{\tilde{\Theta}\right\} > 0$, and a rule c with $\Pr\{c(X) = d|X = x\} > 0$ for some x. By our full support assumption $\Pr_a\left\{c(X) = d|\theta \in \tilde{\Theta}\right\} > 0$, and conditional on $\theta \in \tilde{\Theta}$ the rule c yields strictly higher expected loss than the rule c' which chooses d' whenever c chooses d and agrees with c otherwise. By assumption c' has weakly lower loss for all parameter values $\theta \notin \tilde{\Theta}$, and so dominates c. Hence, any rule admissible in the decision model must choose d with probability zero.

We next show that any rule that chooses d with probability zero is inadmissible in the communication model. Consider any such rule $\tilde{c}$, and for each $\tilde{d} \in \mathcal{D}$ define $\mathcal{X}\left(\tilde{d}\right) = \left\{x \in \mathcal{X} | \Pr\left\{\tilde{c}(X) = \tilde{d}|X = x\right\} > 0\right\}$. If $\cap_{x\in\mathcal{X}(d)} \arg\min_{d\in\mathcal{D}} \mathrm{E}_a[L(d,\theta)|X = x]$ is nonempty for all $d^* \in \mathcal{D} \setminus \{d\}$ and $a \in \mathcal{A}$, then we can show that $N(\mathcal{X},\mathcal{A}) \leq |\mathcal{D}| - 1$. Hence, since $N(\mathcal{X},\mathcal{A}) \geq |\mathcal{D}|$, there exist $d^* \in \mathcal{D} \setminus \{d\}$, $a \in \mathcal{A}$ such that $\cap_{x\in\mathcal{X}(d^*)} \arg\min_{d\in\mathcal{D}} \mathrm{E}_a[L(d,\theta)|X = x] = \emptyset$. For $d^{**} \in \arg\min_{d\in\mathcal{D}} \mathrm{E}_a[L(d,\theta)|\tilde{c}(X) = d^*]$, there exists $\tilde{x} \in \mathcal{X}(d^*)$ and $d^{***} \in \mathcal{D}$ such that

$$\mathrm{E}_a[L(d^{***},\theta)|X = \tilde{x}] < \mathrm{E}_a[L(d^{**},\theta)|X = \tilde{x}].$$

Consider the rule c^* that is equal to $\tilde{c}$ except that it reports d when $X = \tilde{x}$. By Proposition 1, c^* dominates $\tilde{c}$ in communication risk.

Hence, we have shown that any rule admissible in the decision model must choose d with probability zero, while any rule that chooses d with probability zero is inadmissible in the communication model. $\square$

B Extension of Optimal Treatment Assignment Example

This section extends the analysis of optimal treatment assignment in Section 4 to show that when agents have sufficiently informative priors, 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 8. *Suppose that for an audience $\bar{\mathcal{A}}$ and some nonempty set $\mathcal{E} \subseteq \mathcal{X}$,*

$$\arg\max_t \mathrm{E}_a[\theta_t | X] = \arg\max_t \mathrm{E}_a[\theta_t] \text{ for all } a \in \bar{\mathcal{A}}, X \in \mathcal{E}. \tag{2}$$

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

Claim 9. *If in addition to the conditions of Claim 8, (i) $\{X : \arg\max_t X_t = \{1, ..., T\}\} \cap \mathcal{E} \neq \emptyset$, (ii) there exists $a \in \bar{\mathcal{A}}$ with $\arg\max_t \mathrm{E}_a[\theta_t | c^*(X)] = c^*(X)$ for all X, and (iii) $\arg\max_t \mathrm{E}_a[\theta_t]$ is a singleton, then $\bar{c}$ dominates c^* in communication risk.*

Proof of Claim 8. Note that all agents have the option to choose $d \in \arg\max_t \mathrm{E}_a[\theta_t]$ conditional on observing $\bar{c}(X) = \iota$, while choosing $d \in \arg\max_t \mathrm{E}_a[\theta_t | \bar{c}(X)]$ conditional on observing $\bar{c}(X) \neq \iota$. By the definition of $\mathcal{E}$ this yields a weakly lower expected loss for agent a than choosing some $d \in \arg\max_t \mathrm{E}_a[\theta_t | c^*(X)]$. $\square$

Proof of Claim 9. If $\arg\max_t \mathrm{E}_a[\theta_t]$ is a singleton for a given agent a and (2) holds, then conditional on $X \in \mathcal{E}$ agent a strictly prefers not to randomize their decision. At the same time, since $\arg\max_t \mathrm{E}_a[\theta_t | c^*(X)] = c^*(X)$, 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 in the proof of Claim 8, since the agent is free to choose $d = \bar{c}(X)$ conditional on $\bar{c}(X) \neq \iota$ and $d = \arg\max_t \mathrm{E}_a[\theta_t]$ conditional on $\bar{c}(X) = \iota$, we see that $\bar{c}$ yields a strictly lower communication risk for this agent. Since we have shown in the proof of Claim 8 that $\bar{c}$ yields weakly lower communication risk than c^* for all $a \in \bar{\mathcal{A}}$, $\bar{c}$ dominates c^*. $\square$

C Extension of Combining Multiple Moments Example

Building on Section 5.3, now suppose that $X = G\tau + (\eta', \gamma')' + \nu + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2 \cdot I_k)$. The analyst again observes X, while the variance $\sigma^2 > 0$ is commonly known, and the loss is $L(d, \theta) = (d - \tau)^2$. The unknown parameters are $\theta = (\tau, \eta, \gamma, \nu)$. All agents $a \in \mathcal{A}$ have $N(0, \rho_\tau^2)$ priors on τ, dogmatic priors on η with $\Pr_a \{\eta = \eta_a\} = 1$, and $N\left(0, \rho_\gamma^2 \cdot I_{\dim(\gamma)}\right)$ priors on γ independent of τ.

If each agent a believes that $\Pr_a \{\nu = 0\} = 1$, the analysis in this extension coincides with that in Section 5.3. Instead, suppose that each agent a believes that $\nu \sim N(0, V_a)$ for V_a a positive semidefinite matrix, and that ν is independent of (τ, η, γ). To express agent a's posterior mean for τ conditional on X under this assumption, define

$$\Xi_a = \sigma^2 \cdot I_k + \begin{bmatrix} 0 & 0 \\ 0 & \rho_\gamma^2 \cdot I_{\dim(\gamma)} \end{bmatrix} + V_a, \ \sigma_{\tau,a}^2 = \left(G'\Xi_a^{-1}G\right)^{-1}.$$

Agent a's posterior mean for τ (and hence optimal decision) is

$$c^a(X) = \frac{1}{\sigma_{\tau,a}^{-2} + \rho_\tau^{-2}} \left(G'\Xi_a^{-1}\left(X - (\eta_a', 0')'\right)\right).$$

Further suppose that the set of V_a matrices over the audience is given by $\mathcal{V} = \{V_a : a \in \mathcal{A}\} = \zeta \cdot \left\{\text{p.s.d. } \Omega \in \mathbb{R}^{k \times k} : \|\Omega\| \leq 1\right\}$, for $\|\cdot\|$ the Frobenius norm and $\zeta > 0$. Hence, as $\zeta \to 0$, the situation converges to that described in Section 5.3.

We first show that X is a minimal (marginally) sufficient statistic for τ in this setting. Note that since matrix inversion is a homoemorphism between $\{\Xi_a : a \in \mathcal{A}\}$ and $\{\Xi_a^{-1} : a \in \mathcal{A}\}$, $\{\Xi_a^{-1} : a \in \mathcal{A}\}$ has a nonempty interior. This implies, however, that for any $X, \tilde{X} \in \mathbb{R}^k$ with $X \neq \tilde{X}$, there exists $a \in \mathcal{A}$ such that $c^a(X) - c^a\left(\tilde{X}\right) =$

$\frac{G'\Xi_a^{-1}(X-\tilde{X})}{\sigma_{\tau,a}^{-2}+\rho_\tau^{-2}} \neq 0$, i.e., for whom these two realizations of the data imply different optimal decisions.

We next show that the communication risk of $c^*_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)$ as described in Section 5.3 approaches that of the optimal rule based on X as $\zeta \to 0$. Note that Ξ_a^{-1} is continuous in V_a, so as $V_a \to 0$, $\Xi_a^{-1} \to \Xi_0^{-1}$, where $V_0 = 0$, and $c^a(X) \to \mathrm{E}_a[\tau | c^*_\omega(\hat{\tau}_\eta, \hat{\tau}_\gamma)]$ for each realization of X. The dominated convergence theorem thus implies that as $V_a \to 0$, $R^*_a(X) \to R^*_a(c^*_\omega)$, as we aimed to show.

References

Abadie, Alberto. 2020. Statistical non-significance in empirical economics. *American Economic Review: Insights* 2(2): 193-208.

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.

Aliprantis, Charalambos D. and Kim C. Border. 2006. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Berlin: Springer-Verlag.

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. *Journal of Business and Economic Statistics* 38(4): 711-722.

Armstrong, Timothy B. and Michal Kolesár. 2021. Sensitivity analysis using approximate moment condition models. *Quantitative Economics* 12(1): 77-108.

Athey, Susan and Stefan Wager. 2021. Policy learning with observational data. *Econometrica* 89(1): 133-161.

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.

Banerjee, Abhijit, Sylvain Chassang, Sergio Montero, and Erik Snowberg. 2020. A theory of experimenters: robustness, randomization, and balance. *American*

Economic Review 110(4): 1206-1230.

Baranchik, A.J. 1970. A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics* 41(2): 642-645.

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.

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.* Accessed on March 25 2021 at <https://openknowledge.worldbank.org/handle/10986/25820>.

DellaVigna, Stefano. 2018. Structural behavioral economics. In D. Bernheim, S. DellaVigna, and D. Laibson (eds.), *Handbook of Behavioral Economics: Applications and Foundations* 1: 613-723. Elsevier.

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.

Dridi, Ramdan, Alain Guay, and Eric Renault. 2007. Indirect inference and calibration of dynamic stochastic general equilibrium models. *Journal of Econometrics* 136(2): 397-430.

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.

Efron, Bradley and Carl Morris. 1973. Stein's estimation rule and its competitors—an empirical Bayes approach. *Journal of the American Statistical Association* 68(341): 117-130.

Farrell, Joseph and Robert Gibbons. 1989. Cheap talk with two audiences. *American Economic Review* 79(5): 1214-1223.

Frankel, Alex and Maximilian Kasy. Forthcoming. Which findings should be published? *American Economic Journal: Microeconomics.*

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.

Gilboa, Itzhak and David Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2): 141-153.

Hildreth, Clifford. 1963. Bayesian statisticians and remote clients. *Econometrica* 31(3): 422-438.

James, W. and Charles Stein. 1961. Estimation with quadratic loss. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1: 361-379.

Jordan, Michael I., Jason D. Lee, and Yung Yang. 2018. Communication-efficient distributed statistical inference. *Journal of the American Statistical Association* 114(526): 668-681.

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.

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. *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.

Manski, Charles F. 2019. Statistical inference for statistical decisions. *arXiv preprint* arXiv:1909.06853. Accessed on August 25 2020 at <https://arxiv.org/abs/1909.06853>.

Nakamura, Emi and Jón Steinsson. 2018. Identification in macroeconomics. *Journal of Economic Perspectives* 32(3): 59-86.

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. Accessed on March 25, 2021 at <http://sims.princeton.edu/yftp/EmetSoc607/AppliedBayes.pdf>.

Sion, Maurice. 1958. On general minimax theorems. *Pacific Journal of Mathematics* 8(1): 171-176.

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

Strasser, Helmut. 1985. *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory (De Gruyter Studies in Mathematics 7).* Berlin: Walter de Gruyter.

Torgersen, E. 1991. *Comparison of Experiments.* Encyclopedia of Mathematics and its Applications. 1st ed. Cambridge: Cambridge University Press.

von Neumann, John and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior.* Princeton: Princeton University Press.

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

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.

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