The Value of Model Misspecification in Communication*

Simon GLEYZE† Agathe PERNOUD‡

Abstract

How do subjective models of the world affect communication? We introduce a cheap-talk game in which Receiver faces uncertainty on models, i.e., which variables cause some outcome of interest, and uncertainty on states, i.e., the realization of these variables. We show that holding a monocausal model can increase informativeness of equilibrium communication on states. Indeed, monocausal models reduce the number of individually rational actions for Receiver, which limits the extent of information manipulation. Then, we show that a Principal who knows the true model benefits from delegating decision-making to a Receiver who holds a monocausal model, be it misspecified.

Keywords: Strategic Communication, Mental Models, Model Misspecification.

JEL classification codes: C72, D83

*We are grateful to Francis Bloch, Matthew Gentzkow, Matthew Jackson, Philippe Jehiel, Frédéric Koessler, Paul Milgrom, Ronny Razin, Olivier Tercieux for helpful conversations. We also thank seminar participants at PSE, Stanford, and the Akbarpour–Milgrom Monday night group for valuable comments and questions. S. Gleyze acknowledges the support of the EUR grant ANR-17-EURE-0001.

†Uber. Email: gleyze.simon@gmail.com

‡Department of Economics, Stanford University. Email: agathep@stanford.edu

1 INTRODUCTION

Why are people drawn to monocausal explanations of complex social phenomena? One of the distinctive features of the increasingly popular populist narratives is their simplicity: many complex problems boil down to a unique explanation, such as immigration, the welfare state, or bureaucracy.[1] For instance, immigration is very often blamed for increasing unemployment, diverting public spending away from citizens, and provoking a cultural war. These narratives, however, seem to overlook many important and relevant factors at play. Is this misspecification the result of bounded rationality, incomplete information or can it have instrumental value?

In this paper, we argue that model misspecification can have *instrumental value* because it acts as a commitment device in strategic communication games between a Receiver and a Sender. We introduce a two-dimensional cheap talk game in which Receiver faces two types of uncertainty. First, she does not know which variables are payoff-relevant—this is referred to as "model uncertainty". Second, Receiver does not know the realization of these variables—which is the more standard "state uncertainty". Receiver communicates with a Sender who is informed about the state. To fix ideas, let Receiver be a politician (e.g., the President) and Sender be a lobbyist or an advisor.[2] The politician would like to approve policies that address some economic problem of interest. The success of each policy depends on which variables are the cause of the problem (the true "model") and their realizations (the realized state). The lobbyist only has partially aligned preferences with the politician.

We first take as given Receiver's worldview[3] and study how it affects communication with Sender. Our first main result is that holding a misspecified worldview can increase the informativeness of equilibrium communication between Sender and Receiver. The intuition is that monocausal models—i.e., believing that few variables are actually relevant—reduce the number of individ-

[1] The idea that populism is characterized by the simplicity of its narratives received empirical support in Bischof and Senninger (2018) who studies political manifestos in Austria and Germany between 1945 and 2013.

[2] We also consider an application in which Receiver is a CEO and Sender is a syndicate.

[3] A *model* specifies which variables cause the outcome, whereas a *worldview* is a belief over models.

ually rational actions for Receiver, which can lead to more communication as there is less room for information manipulation. We then show that holding such a model, even if it is misspecified, can be welfare improving for Receiver.

We then endogenize Receiver's worldview by introducing another agent, the Principal (e.g. a voter), who is informed about the true model of the world. If the Principal can delegate decision-making to a Receiver with a specific worldview, we show that he always prefers a Receiver with a worldview that puts significant weight on a simple, monocausal model. Hence model misspecification, defined as Receiver holding an incorrect worldview given the true realized model, can have instrumental value in strategic communication. Framed in the context of our running example, this suggests voters may elect populist politicians not because they agree with their worldviews, but because they believe such candidates will not be "pushovers" that are easily influenced by lobbyists, their advisors or the administration.[4] Similarly, if the Principal can only communicate on models (so he cannot directly choose the worldview of Receiver, but can try to influence it via communication), we show that all equilibria are outcome-equivalent to a babbling equilibrium—namely, communication on models is impossible. This is precisely driven by the fact that the Principal benefits from Receiver holding a misspecificed worldview, which prevents meaningful communication.

Related Literature We contribute to the literature on (multidimensional) cheap talk, building on the model of Crawford and Sobel (1982). Under the assumption of state independent preferences, Chakraborty and Harbaugh (2010) extend the basic model to a multi-dimensional setting and prove equilibrium existence. Lipnowski and Ravid (2020) further provide a geometric characterization of equilibrium payoffs. Levy and Razin (2007) show that the correlation structure between dimensions of the state puts bounds on equilibrium communication. The main innovation of our paper is to introduce subjective models—describing which variables Receiver should act on—and study how

[4]Most common explanations of populism borrow concepts from identity politics, and range from rising inequalities to cultural backlash (see Guriev and Papaioannou (2022) for an extensive review). To the best of our knowledge, there is no investigation of the instrumental value of populist ideology to date.

they affect communication. Of course, what we call “models” could be embedded in a larger state space, but we think that the conceptual distinction is relevant and allows to study the formation of worldviews. Related to this literature on cheap talk and the present paper, Che and Kartik (2009) show that in a disclosure game the Receiver may choose a Sender with a different prior to incentivize information acquisition.

Schwartzstein and Sunderam (2021) introduce the idea of communication on models by having Receiver adopt Sender’s proposed model whenever it explains past data better than his default model. Receiver has no prior on models and communication is not strategic, whereas in our model Receiver is Bayesian and communication is strategic. This leads to different equilibrium predictions, as communication on models is impossible in our setting, whereas in Schwartzstein and Sunderam (2021) it is easy to change Receiver’s worldview even when the default model is correct.

We also contribute to the literature studying the value and implications of model misspecification. Eliaz and Spiegler (2020) show that agents maximizing anticipatory utility tend to choose misspecified models. Olea et al. (2022) show that agents that have simple models, i.e. use few variables to predict an outcome, have more confidence in their estimate when sample size is small. Levy et al. (2022) introduce a model of political competition between two groups—one has a simpler subjective model than the other—and show that this leads to policy cycles and extreme policy choices. Instead, we study how misspecified models affect strategic communication.

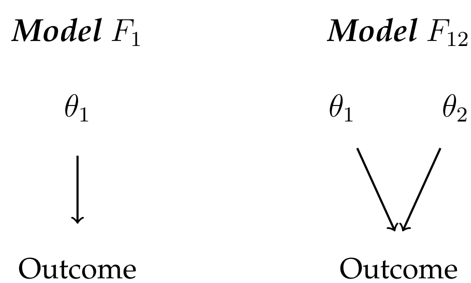
Finally, observe that our concept of worldview can serve as a commitment device on individually rational actions. Therefore, we speak to the literature studying Receiver’s commitment power in strategic communication. A key result is that Receiver often wants to commit to underreact to the information provided by Sender (Whitmeyer (2019); Ball (2019); Frankel and Kartik (2022)), e.g., by hiring an intermediary that garbles said information. In our paper, holding a monocausal worldview also allows Receiver to “underreact” to Sender’s information—hence the intuition is similar, but the mechanism at play and its implications are different. We further discuss the relation between our model and one in which Receiver has commitment power in Section 4.

2 How Models Shape Communication

2.1 Setup

We consider a cheap talk game with two types of uncertainty. Receiver faces "model uncertainty": she does not know which variables, i.e., which dimensions of the state space, are payoff-relevant. Moreover, Receiver also faces the traditional "state uncertainty", i.e. uncertainty about the realization of these variables. In this section, we consider the case of a unique Sender, who is informed about the state and can communicate on it.

Let $\Theta = \{\underline{\theta}_1, \bar{\theta}_1\} \times \{\underline{\theta}_2, \bar{\theta}_2\} \equiv \{0,1\}^2$ be the state space, so that the state is composed of two variables—θ_1 and θ_2—each of which can either be high or low. A *model* specifies which of these variables cause an economic outcome of interest.[5] Formally, models can be thought of as directed acyclic graphs (DAGs) in which a variable is an ancestor of another in the graph if it is one of its cause. There are two possible models:[6]



According to Model F_1, only variable θ_1 can cause the outcome of interest, whereas under Model F_{12} both variables can. The true model is unknown to Receiver. Receiver believes that the true model is F_1 with probability λ, and F_{12} with complementary probability. This section investigates how Receiver's *worldview*, i.e., the relative weight λ she puts on Model F_1, affects equilibrium communication.

[5] In reality, Receiver probably has uncertainty on the entire joint distribution of variables, but as a first step we only consider uncertainty on *which* variables she should care about.

[6] We exclude the two remaining DAGs (the one in which only θ_2 causes the outcome and the one in which neither θ_1 nor θ_2 does) from our main analysis. In Section 4, we show that this is without loss as only entertaining Models F_1 and F_{12} is payoff-maximizing for Receiver.

Receiver can take action to address each of the two possible causes. Let $A = \{\underline{a}_1, \overline{a}_1\} \times \{\underline{a}_2, \overline{a}_2\} \equiv \{0,1\}^2$ be the action space. Interpret the high action $\overline{a}_k$ as Receiver taking active measures to reduce variable θ_k, and the low action $\underline{a}_k$ as the status-quo. Receiver only wants to act on the true causes of the problem. If the true model is F_1, the optimal action on the first dimension is $a_1^*(\theta_1, F_1) = \overline{a}_1$ if $\theta_1 = \overline{\theta}_1$ and $a_1^*(\theta_1, F_1) = \underline{a}_1$ if $\theta_1 = \underline{\theta}_1$. It is however optimal to set $a_2^*(\theta_2, F_1) = \underline{a}_2$ irrespective of θ_2 as that variable does not contribute to the outcome under model F_1. It is *as if* high actions were costly—hence if a variable is irrelevant then the status-quo is optimal. On the contrary, if the true model is F_{12}, then the optimal action along both dimensions is to match the state.

Instead of defining Receiver's preferences on final outcomes, we directly define preferences on actions, states and models. We show in Section 4 how these reduced-form preferences can be microfounded using final outcomes. Receiver's payoff from action (a_1, a_2) in state (θ_1, θ_2) if the true model is F writes:[7]

$$u_R(a_1, a_2, \theta_1, \theta_2, F) = -(a_1 - a_1^*(\theta_1, F))^2 - (a_2 - a_2^*(\theta_2, F))^2.$$

Sender's preferences are only partially aligned with Receiver's:

$$u_S(a_1, a_2, \theta_1, \theta_2, F) = -(a_1 - a_1^*(\theta_1, F))^2 + \gamma a_2$$

with $\gamma > 1$. Namely, Sender and Receiver are aligned on the first issue θ_1 but Sender has an agenda on the second issue θ_2 and wants higher action regardless of the true model or the realized state. This misalignment might prevent full communication about the state in equilibrium, and is key for our analysis.

The joint distribution of states is

	$\theta_2 = 0$	$\theta_2 = 1$
$\theta_1 = 0$	$1 - \mu_0$	0
$\theta_1 = 1$	0	μ_0

with $\mu_0 < 0.5$. For tractability purposes, we restrict attention to the case of

[7] Assuming quadratic preferences for Receiver simplifies the exposition greatly, but does not seem to be driving our results. Most of them extend when the gains from taking the correct action depend on the dimension, the state, and the model.

perfect correlation between the two variables. We discuss at the end of the paper how our results generalize to settings with negative correlation.

The timing of the game is as follows: First, Sender observes the state and sends a message $m \in M$ to Receiver, where M is a finite set with $|M| \geq 2$. Sending messages is free, and Sender cannot commit to a specific communication protocol ex ante (*cheap talk*). Receiver then takes an action. A (perfect Bayesian) equilibrium consists of a strategy $q_S : \Theta \longrightarrow \Delta M$ for Sender, a strategy $p_R : M \longrightarrow \Delta A$ for Receiver, and a belief system such that (i) Receiver's beliefs are derived from the prior μ_0 and q_S using Bayes' rule whenever possible, (ii) Receiver only plays actions that are optimal given her belief, and (iii) Sender only sends messages that maximize his expected utility given θ. Receiver's worldview λ is taken as given and is known to Sender. We focus on the most informative equilibrium, which is the one preferred by all agents.

Example 1 (Political Economy): Receiver is an elected politician who decides on policies (e.g., the President). Sender is a lobbyist or an advisor who communicates on the state of the economy. The outcome of interest is unemployment, which has two possible causes: the extent of immigration θ_1 and of aggregate consumption θ_2. The politician aims at reducing unemployment, and to that end wants to address whichever variable(s) cause(s) it. To address immigration, she can impose additional legal requirements for new entrants in the country ($a_1 = 1$). To address low aggregate consumption, she can undertake an expansionary fiscal policy ($a_2 = 1$). The lobbyist wants greater public spending regardless of whether it is sufficiently high already, and whether low aggregate consumption is actually contributing to unemployment. Because of this agenda, the lobbyist tries to influence policy by exaggerating how low aggregate demand is.

Example 2 (Organization): Receiver is an employer who manages a firm (e.g., the CEO). Sender is an employee, or a worker syndicate, who communicates on issues faced by workers in the production process. The outcome of interest is (low) productivity, which has two possible causes: skill mismatch between workers and the task θ_1 and workers' effort θ_2. The CEO aims at increasing the firm's productivity, and to that end wants to address whichever variable(s) depress(es) it. To address skill mismatch, she can provide additional training to the workers ($a_1 = 1$). To incentivize higher effort, she can increase wages ($a_2 = 1$). The worker syndicate wants higher wages regardless of whether they are sufficiently high already, or whether they are at all related to productivity. Because of this agenda, the syndicate tries to influence the CEO's decisions by exaggerating how costly effort is for workers.

2.2 Equilibrium Characterization

Let μ be the posterior probability that Receiver assigns to state $\theta = (1,1)$. Receiver's optimal action as a function of μ and λ is

$$\sigma^*(\mu, \lambda) = \begin{cases} (0,0) & \text{if } \mu \leq \frac{1}{2} \\ (1,0) & \text{if } \mu \geq \frac{1}{2} \text{ and } \mu \leq \mu^* \equiv \frac{1}{2(1-\lambda)} \\ (1,1) & \text{if } \mu \geq \mu^* \end{cases}$$

Because the first variable (θ_1) is causing the outcome under both model F_1 and F_{12}, whether or not Receiver wants to act on it only depends on her belief about the state μ. On the contrary, the second variable (θ_2) only contributes to the outcome under model F_{12}. Hence Receiver wants to take a high action $a_2 = 1$ only if she puts sufficient weight on both the high state *and* model F_{12}.

First, note that for $\lambda > 0.5 \equiv \lambda^*$, Receiver never takes action $a_2 = 1$. Hence the only actions that Sender can induce are $a = (0,0)$ and $a = (1,0)$, and since his preferences are aligned with Receiver's over those, the most informative equilibrium is fully revealing.

Second, for $\lambda \leq \lambda^*$ it is possible to induce action $a_2 = 1$. What is not possible, however, is that in equilibrium Sender sends a message m that induces actions $a = (1,1)$ with probability one. Sending that message would ensure Sender a payoff of

$$-1 + \gamma > 0$$

in state $\theta = (0,0)$, and Sender would *always* want to send it. Hence Receiver's posterior must always lie weakly below μ^* in equilibrium.

The most informative equilibrium has then Sender send two messages m_1 and m_0.[8] Message m_0 indicates the low state with certainty, and $\Pr(a = (0,0) \mid m_0) = 1$. Message m_1 leads Receiver to randomize between actions $(1,1)$ and $(1,0)$. In state $\theta = (0,0)$, Sender must then be indifferent between inducing action $(0,0)$ for sure and inducing a lottery over actions $(1,1)$ and $(1,0)$, which requires

$$\begin{aligned}
&\Pr(a = (1,1) \mid m_1)\mathbb{E}_\lambda[u_S(\overline{a}_1, \overline{a}_2, \underline{\theta}_1, \underline{\theta}_2, F)] \\
&\qquad + \Pr(a = (1,0) \mid m_1)\mathbb{E}_\lambda[u_S(\overline{a}_1, \underline{a}_2, \underline{\theta}_1, \underline{\theta}_2, F)] = \mathbb{E}_\lambda[u_S(\underline{a}_1, \underline{a}_2, \underline{\theta}_1, \underline{\theta}_2, F)] \\
\iff &\Pr(a = (1,1) \mid m_1)[-1 + \gamma] - \Pr(a = (1,0) \mid m_1) = 0.
\end{aligned}$$

Namely,

$$\Pr(a = (1,1) \mid m_1) = \frac{1}{\gamma}.$$

Since Receiver must be indifferent between taking the low and high action a_2 upon receiving m_1, her belief must be equal to μ^*. Sender's strategy must then satisfy:

$$\Pr(\theta = (1,1) \mid m_1) = \frac{\Pr(m_1 \mid \theta = (1,1))\mu_0}{\Pr(m_1 \mid \theta = (1,1))\mu_0 + \Pr(m_1 \mid \theta = (0,0))(1 - \mu_0)} = \mu^*.$$

Overall, the communication strategy of Sender is

$$q_S(m_1 \mid \theta = (1,1)) = 1, \qquad q_S(m_1 \mid \theta = (0,0)) = \frac{\mu_0}{1 - \mu_0}\frac{1 - \mu^*}{\mu^*}$$

and m_0 with complementary probability. In response, Receiver chooses the following distribution over actions

$$p_R(a = (0,0) \mid m_0) = 1, \quad p_R(a = (1,1) \mid m_1) = \frac{1}{\gamma}, \quad p_R(a = (1,0) \mid m_1) = 1 - \frac{1}{\gamma}.$$

Our first main result is that holding a simpler worldview (i.e., putting more

[8] It is actually the only equilibrium that is not outcome-equivalent to a babbling equilibrium as we show in the Appendix.

weight on the single-cause model F_1) makes the equilibrium more informative.

PROPOSITION 1. *The informativeness (in the Blackwell sense) of the equilibrium is monotonically increasing in λ.*

This follows directly from the above characterization of equilibrium behavior: the probability that Sender sends the "wrong" message m_1 in state $\theta = (0, 0)$ decreases in λ. Indeed, if Receiver puts a lot of weight on the simple model F_1, then she needs to be almost certain that the state is high to be willing to take action $a_2 = \overline{a}_2$. Hence more communication is required.

Next, we show that Receiver benefits from putting more weight on the monocausal model F_1. Let $V(\lambda)$ denote the expected utility of Receiver in equilibrium, as a function of the probability she assigns to model F_1:

$$V(\lambda) \equiv \sum_{\theta} \mu_0(\theta) \sum_{m} q_S(m|\theta; \lambda) \sum_{a} p_R(a|m; \lambda) \mathbb{E}_{\lambda}[u_R(a, \theta, F)].$$

PROPOSITION 2. *The expected utility of Receiver $V(\lambda)$ is monotonically increasing in λ.*

Note that this is not a straightforward implication of Blackwell's theorem as Receiver's worldview λ has two effects: it impacts equilibrium play—q_S and p_R,—as well as how Receiver evaluates the outcome induced by equilibrium play—$\mathbb{E}_{\lambda}[u_R(\cdot)]$. As λ increases, the equilibrium becomes more informative (Proposition 1) and allows Receiver to better target action a_1 to the realized state θ_1. The same is however not true for action a_2, as when λ goes above λ^* Receiver stops taking action $a_2 = \overline{a}_2$ altogether, which is costly if the true model is F_{12}. Hence Receiver trades-off better decision-making on the first dimension, with potentially more mistakes on the second. Since a greater λ also means that Receiver puts more weight on model F_1 when evaluating the equilibrium outcome, not taking action $a_2 = \overline{a}_2$ is less likely to be a mistake, and her overall expected payoff is larger.

3 The Value of Model Misspecification

In the previous section, we showed that Receiver's expected utility is increasing in her subjective belief in the simple model F_1. This is due to two effects: a strategic effect (informativeness of communication increases), and a preference effect (Receiver believes she is doing fewer mistakes). In this section, we disentangle them, and show that even if we neutralize the preference effect by looking through the lens of an informed Principal, the Principal is still better off with Receiver holding a simpler misspecified model. Therefore, this shows that misspecification can have a positive value in strategic communication.

3.1 Delegation

For the following two sections, we extend our framework and introduce another agent, a Principal. The Principal is informed of the true model F, but does not know the state θ. Instead of communicating directly with Sender and making decisions himself, the Principal can delegate decision-making to an agent, who is the Receiver from the previous section. The latter then communicates on states with Sender in a second stage. Receiver and Principal share the same preferences, but not the same worldview: Principal knows the true model whereas Receiver puts weight λ on model F_1.

Let $V_P(\lambda|F)$ denote the expected equilibrium payoff of Principal when the true model is F and he delegates decision-making to a Receiver with worldview λ. Using the above characterization of equilibrium communication, these write

$$V_P(\lambda|F_{12}) \equiv \sum_{\theta} \mu_0(\theta) \sum_{m} q_S(m|\theta;\lambda) \sum_{a} p_R(a|m;\lambda) u_R(a,\theta,F_{12})$$
$$V_P(\lambda|F_1) \equiv \sum_{\theta} \mu_0(\theta) \sum_{m} q_S(m|\theta;\lambda) \sum_{a} p_R(a|m;\lambda) u_R(a,\theta,F_1)$$

Figure 1 plots these expected payoffs.

We assume Principal can choose the worldview of Receiver—perhaps because there is a vast pool of agents with various worldviews from which Principal can hire. Say Receiver's worldview is misspecificed if her belief about

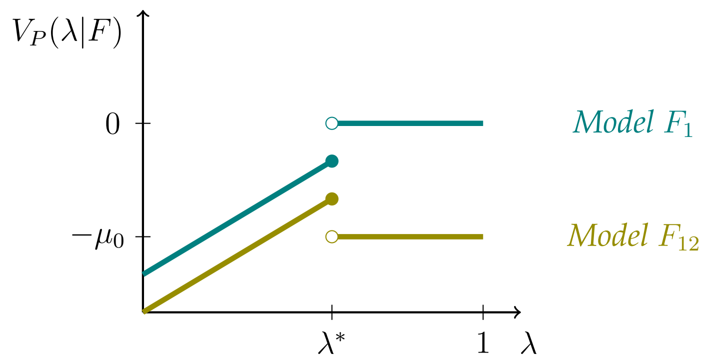


Figure 1: *Principal's equilibrium expected payoff as a function of Receiver's worldview λ, for each possible realization of the true model.*

models is incorrect, that is if $\lambda > 0$ while the true model is F_{12}, or if $\lambda < 1$ while the true model is F_1.

Proposition 3. *When the true Model is F_1, choosing a Receiver with the correct worldview $\lambda = 1$ is optimal for the Principal.*[9] *When the true Model is F_{12}, the Principal optimally chooses a Receiver with a misspecified, simpler worldview:*

$$\arg\max_{\lambda} V_P(\lambda|F_{12}) = \lambda^* > 0.$$

Therefore, even when Principal knows that the true model is complex and multi-causal F_{12}, he chooses a Receiver who puts significant weight on the simpler, unicausal model.[10] Holding such misspecified worldview is beneficial, as it attenuates how much Sender tries to mislead Receiver to serve his own agenda. This is true even though a misspecified worldview leads Receiver to take actions that are ex post suboptimal in the eyes of Principal, as it makes Receiver "too conservative" when it comes to taking action $a_2 = \overline{a}_2$.

In example 1, the Principal can be thought of as a voter who chooses a populist representative in the hope that he will not get fooled by lobbyists or advisors. Proposition 3 says that voting for such populist representative is optimal even when the voter does not share the populist's worldview, but instead be-

[9] Note that all worldviews $\lambda > 0.5$ are optimal as they all lead to the same (fully revealing) equilibrium.

[10] If Receiver can only hold a deterministic worldview (i.e., $\lambda \in \{0, 1\}$), then Principal optimally chooses a Receiver fully believing in the *wrong* model ($\lambda = 1$) as $V_P(0|F_{12}) < V_P(1|F_{12})$.

lieves in a more complex, multi-causal model of the world. In example 2, the Principal can be thought of as a board of directors who chooses a conservative CEO who is skeptical about low wages ever contributing to low productivity, and will not be easily manipulated by syndicates. Again, choosing such a CEO is optimal even when the board of directors does not share the CEO's world-view.

3.2 Communication on Models

Now suppose that Principal is not able to directly *choose* the worldview of Receiver—perhaps the pool of agents is not that rich, or an agent's worldview is hard to identify at the time of the hire. What Principal can however do is communicate about the true model with Receiver. For instance, in Example 2, the board of directors can share with the CEO their accumulated knowledge of the workings of the firm, and in particular of how much effort and efficient wage levels affect its overall productivity.

The timing of the game is as follows. First, the true model is drawn according to a prior distribution $\lambda_0 = \Pr(F_1)$ and is observed by the Principal. Principal sends a message $n \in N$ from some arbitrary set of messages. To keep the analysis simple and uncluttered, suppose that this message is public, in the sense that it is observed by both Receiver and State-Sender. In a second stage, the state is drawn and observed by State-Sender. The game then unfolds as before. The previous analysis hence characterizes what happens in this second stage of communication, given some belief λ that resulted from communication with Principal. We solve for the equilibrium communication on models in the first stage.

As before, the expected payoff of Principal when Receiver holds posterior λ while the realized model is F is $V_P(\lambda|F)$. Let $q_P : \{F_1, F_{12}\} \longrightarrow \Delta N$ denote Principal's communication strategy. Receiver's belief about models upon receiving n is derived from Bayes' law:

$$\lambda(n) = \frac{q_P(n|F_1)\lambda_0}{q_P(n|F_1)\lambda_0 + q_P(n|F_{12})(1 - \lambda_0)}$$

for all $n \in \operatorname{supp} q_P \equiv \{n | q_P(n|F) > 0 \text{ for some } F\}$. For q_P to be an equilibrium, we must have that for all F, $q_P(\cdot|F)$ is supported on

$$\arg\max_{n \in N} V_P(\lambda(n)|F).$$

If not, Principal must sometimes be sending a strictly suboptimal message for some realization of the model F, and must hence have an incentive to deviate.

Can Principal provide meaningful information about the realized model to Receiver? We show that the answer is no: Principal benefits from Receiver holding a misspecified model, which prevents credible communication.

PROPOSITION 4. *In any equilibrium, communication on models in the first stage is payoff-equivalent to a babbling equilibrium.*

Figure 1 is helpful to understand why communication is impossible. Any informative equilibrium must involve at least one message n_{12} that indicates model F_{12} is more likely—namely, $\lambda(n_{12}) < \lambda_0$—and another m_1 that indicates the opposite—$\lambda(n_1) > \lambda_0$. If the true model is F_1, Principal has a strict incentive to be truthful as soon as $\lambda(n_{12}) \leq \lambda^*$.[11] But then Receiver knows that message n_{12} can never be sent when $F = F_1$, and so his posterior worldview must be $\lambda(n_{12}) = 0$. This is however the payoff-*minimizing* worldview under both realizations of the model, as Receiver benefits from putting some weight on model F_1 even when the true model is F_{12}. Hence, upon observing F_{12}, Principal has a strict incentive to deviate and send whichever message indicates F_1 is more likely. Despite the facts that Principal's preferences are fully aligned with Receiver's, communication is very limited so as to prevent Receiver from being manipulated in the second stage by State-Sender.

4 DISCUSSION

Misspecified Models vs. Commitment Power. Holding a misspecified model of the world can be valuable in strategic communication as it allows Receiver

[11] If $\lambda(n_{12}) > \lambda^*$, i.e., Receiver's posterior always lies above λ^*, then equilibrium communication with State-Sender is unaffected by first-stage communication. Indeed, for all $\lambda > \lambda^*$, State-Sender fully reveals the state in equilibrium and Receiver never takes action $a_2 = 1$.

to be more conservative when it comes to taking actions towards which Sender is biased. One can think of this as a form of commitment power: if Receiver puts more weight on model F_1, she needs more evidence that the high state $\theta = (1,1)$ has realized to be willing to take action $a_2 = \overline{a}_2$. Of course, worldviews give Receiver much less flexibility than if she could fully commit to a decision rule $\rho : M \longrightarrow \Delta A$, as Receiver's actions must still be consistent with her posterior about the state. Yet, we show that, in our setting, the optimal decision rule under full commitment can be implemented by holding a specific worldview, and behaving optimally according to that worldview.

PROPOSITION 5. *Suppose the true model is F_{12}.*[12] *When she holds the optimal misspecified worldview $\lambda = \lambda^*$, Receiver achieves the same expected equilibrium payoff as if she had full commitment power.*

If Receiver could fully commit to a decision rule, then without loss she would incentivize Sender to fully reveal the state, and would take the optimal action as often as possible while respecting this incentive compatibility constraint. To that end, she would commit to mix between $a = (1, 0)$ and $a = (1, 1)$ when Sender tells her the state is high, to ensure Sender remains truthful when the state is low. This coincides precisely with the equilibrium derived above when Receiver puts weight $\lambda = \lambda^*$ on model F_1. Hence one can think of a worldview as one way of implementing the optimal decision rule.

Correlation Across States. All our analysis extends to the case of perfect negative correlation, i.e. $\Pr(\theta = (0, 1)) = \mu_0$ and $\Pr(\theta = (1, 0)) = 1-\mu_0$. More generally, the analysis extends to any setting with sufficiently high correlation across states. This is due to the fact that when correlation is high, the equilibrium is constrained to be a monotonic partition, i.e. there is a message that indicates low states and another that indicates high states. Hence, increasing λ makes the incentive constraint tighter and necessarily leads to more information revelation. When variables are independent, however, there exist equilibria that do

[12] When the true model is F_1, commitment power has no value. Indeed, Receiver then never acts on the second issue and always sets $a_2 = 0$. Since Sender and Receiver share the same preferences over the first issue, the equilibrium is fully revealing and cannot be improved upon by commitment power.

not have this monotonic structure. Hence, increasing λ do not necessarily lead to more information revelation. This means that while our result is a proof of principle that holding a misspecified model can have value in strategic communication, this need not extend to all environments. For the intuition behind our results to go through, there must be something linking actions on the two dimensions together—e.g., correlation between θ_1 and θ_2.

Alternative Models. In our main analysis, Receiver only entertains two possible models: F_{12} and F_1. We now investigate what happens under the more general set of possible models $\mathcal{F} \equiv 2^{\{1,2\}}$, where a model $F \in \mathcal{F}$ specifies which variables, if any, cause the outcome.

PROPOSITION 1′. *Let $\lambda, \lambda' \in \Delta\mathcal{F}$ be be two possible worldviews, with $\lambda'(2 \in F) < \lambda(2 \in F)$ and $\lambda'(1 \in F) = \lambda(1 \in F)$. The equilibrium is more informative (in the Blackwell sense) when Receiver holds worldview λ' than when she holds worldview λ.*

Putting more weight on models that exclude variable θ_2 as a potential cause of the outcome makes equilibrium communication more informative. The same is not true of models that exclude variable θ_1. Hence entertaining a monocausal worldview helps only if it reduces the effective misalignement between Sender and Receiver, by ruling out dimensions on which their preferences conflict. The payoff-maximizing worldview characterized in Section 3 is then still optimal in this more general setup.

PROPOSITION 3′. *Suppose the true Model is F_{12}. Choosing a Receiver with a worldview that puts weight λ^* on Model F_1 and remaining weight $1 - \lambda^*$ on Model F_{12} is payoff-maximizing for the Principal.*

The newly introduced models cannot reduce the preference misalignment between Sender and Receiver, and so only entertaining models F_1 and F_{12} is optimal.

Micro-Foundation of Preferences. Let $y \in Y \subseteq \mathbb{R}$ denote the outcome variable of interest. Let $c_k > 0$ denote the cost associated with taking action $a_k = \overline{a}_k$. A model specifies a data generating process, that is how the distribution of y

depends on the realized state and the action taken by Receiver. Formally, there exist model-dependent probability measures $\Pr(\cdot \mid F)$ over $Y \times \Theta \times A$. If a variable does not belong to the model $k \notin F$, then it is as if the variable θ_k were not causally related to y, such that the distribution of y is independent of dimension k: $\Pr(y \mid \theta, a; F) = \Pr(y \mid \theta_{-k}, a_{-k}; F)$. If a variable belongs to the model $k \in F$, then it is a cause of the outcome y and impacts its distribution.

Receiver has two possible actions associated with each variable k: a low action $\underline{a}_k$, which is costless; and a high action $\overline{a}_k$, which costs c_k. Taking the low action should be interpreted as maintaining the status-quo, as opposed to actively addressing variable k, which requires time and effort. When all relevant variables are low—i.e., $\theta_k = \underline{\theta}_k$ for all $k \in F$—the outcome is as high as possible. In this case, actions have no impact: $\mathbb{E}[y \mid \underline{\theta}, a; F] = 0$ for all $a \in A$. A high realization of a payoff-relevant variable $k \in F$ induces the following expected outcome: $\mathbb{E}[y \mid \overline{\theta}, \underline{a}; F] < 0 = \mathbb{E}[y \mid \overline{\theta}, \overline{a}; F]$. Overall, the expected value of the outcome variable of interest can be written as

$$\mathbb{E}[y \mid \theta, a; F] = -\sum_{k \in F} \mathbb{1}\{\theta_k = \overline{\theta}_k, a_k = \underline{a}_k\}.$$

Receiver's payoff equals this expected outcome net of action costs:

$$u_R(a, \theta, F) = \mathbb{E}[y \mid \theta, a; F] - \sum_{k=1,2} c_k \mathbb{1}\{a_k = \overline{a}_k\}.$$

The reduced-form preferences that we use in the main analysis are equivalent to the above preferences with $c_k = 0.5$ for $k = 1, 2$.

APPENDIX A PROOFS

Preliminary Analysis. We show that the equilibrium derived in the body of the text is the only equilibrium that is not outcome-equivalent to babbling. Take any equilibrium and let M^* the set of messages sent with positive probability. As argued above, it cannot be that $\mu(m) > \mu^*$ for some $m \in M^*$. Furthermore, it cannot be that $\mu(m) < \mu^*$ for all $m \in M^*$. Indeed, that would mean Receiver's action only varies over the first dimension across messages, and Sender would always want to induce the highest (resp. lowest) action in the high (resp. low) state, hence fully revealing the state. Overall, $\mu(m) = \mu^*$ for some m.

If $|M^*| = 2$, then the other message must induce a posterior strictly below 0.5 by Bayes' rule. Hence that other message leads Receiver to take action $(0,0)$ and can never be sent by Sender when the $\theta = (1,1)$. We are back to the equilibrium we focus on in our analysis.

If $|M^*| > 2$, then by the previous argument there must exist $m_0 \in M^*$ that leads to $\mu(m_0) = 0$ and $p_R(m_0) = (0,0)$. Any other, non-redundant, message must lead to a posterior in $[0.5, \mu^*)$ and induce action $(1,0)$. Such message would however never be sent when the $\theta = (0,0)$, hence inducing a belief $\mu = 1$, which is impossible.

Proof of Proposition 2. Receiver's expected payoff in equilibrium equals

$$\begin{aligned} V(\lambda) = \mathbf{1}\{\lambda \leq \lambda^*\} \Big[&\mu_0\Big(\frac{1}{\gamma}\mathbb{E}_\lambda[u(\overline{a}_1, \overline{a}_2, \overline{\theta}_1, \overline{\theta}_2, F)] + \frac{\gamma-1}{\gamma}\mathbb{E}_\lambda[u(\overline{a}_1, \underline{a}_2, \overline{\theta}_1, \overline{\theta}_2, F)]\Big) \\ &+(1-\mu_0)\frac{\mu_0}{1-\mu_0}\frac{1-\mu^*}{\mu^*}\Big(\frac{1}{\gamma}\mathbb{E}_\lambda[u(\overline{a}_1, \overline{a}_2, \underline{\theta}_1, \underline{\theta}_2, F)] + \frac{\gamma-1}{\gamma}\mathbb{E}_\lambda[u(\overline{a}_1, \underline{a}_2, \underline{\theta}_1, \underline{\theta}_2, F)]\Big) \\ &+(1-\mu_0)\left(1 - \frac{\mu_0}{1-\mu_0}\frac{1-\mu^*}{\mu^*}\right)\mathbb{E}_\lambda[u(\underline{a}_1, \underline{a}_2, \underline{\theta}_1, \underline{\theta}_2, F)]\Big] \\ &+\mathbf{1}\{\lambda > \lambda^*\}\Big[\mu_0\mathbb{E}_\lambda[u(\overline{a}_1, \underline{a}_2, \overline{\theta}_1, \overline{\theta}_2, F)] + (1-\mu_0)\mathbb{E}_\lambda[u(\underline{a}_1, \underline{a}_2, \underline{\theta}_1, \underline{\theta}_2, F)]\Big]. \end{aligned}$$

Using the fact that $\mu^* = \frac{1}{2(1-\lambda)}$, this simplifies to

$$V(\lambda) = -\mathbf{1}\{\lambda \leq \lambda^*\}\mu_0(2-3\lambda) - \mathbf{1}\{\lambda > \lambda^*\}\mu_0(1-\lambda).$$

It is then easily verified that $V(\lambda)$ is monotonically increasing over $[0, \lambda^*]$ and $(\lambda^*, 1]$, and that it is continuous at $\lambda^* = 0.5$.

□

Proof of Proposition 3. We derive the expected payoff of Receiver in equilibrium given λ, for each possible realization of the true model:

$$V(\lambda|F_{12}) = \mathbf{1}\{\lambda \leq \lambda^*\}\Big[\mu_0\Big(\frac{1}{\gamma}u(\overline{a}_1, \overline{a}_2, \overline{\theta}_1, \overline{\theta}_2, F_{12}) + \frac{\gamma-1}{\gamma}u(\overline{a}_1, \underline{a}_2, \overline{\theta}_1, \overline{\theta}_2, F_{12})\Big)$$

$$
\begin{aligned}
&+(1-\mu_0)\frac{\mu_0}{1-\mu_0}\frac{1-\mu^*}{\mu^*}\Big(\frac{1}{\gamma}u(\overline{a}_1,\overline{a}_2,\underline{\theta}_1,\underline{\theta}_2,F_{12})+\frac{\gamma-1}{\gamma}u(\overline{a}_1,\underline{a}_2,\underline{\theta}_1,\underline{\theta}_2,F_{12})\Big)\\
&+(1-\mu_0)\left(1-\frac{\mu_0}{1-\mu_0}\frac{1-\mu^*}{\mu^*}\right)u(\underline{a}_1,\underline{a}_2,\underline{\theta}_1,\underline{\theta}_2,F_{12})\Big]\\
&+\mathbf{1}\{\lambda>\lambda^*\}\Big[\mu_0 u(\overline{a}_1,\underline{a}_2,\overline{\theta}_1,\overline{\theta}_2,F_{12})+(1-\mu_0)u(\underline{a}_1,\underline{a}_2,\underline{\theta}_1,\underline{\theta}_2,F_{12})\Big]\\
&=-\mathbf{1}\{\lambda\le\lambda^*\}\frac{2\mu_0}{\gamma}[\gamma-(1+\gamma)\lambda]-\mathbf{1}\{\lambda>\lambda^*\}\mu_0.
\end{aligned}
$$

$$
V(\lambda|F_1)=-\mathbf{1}\{\lambda\le\lambda^*\}\frac{\mu_0}{\gamma}[1+(1-2\lambda)(\gamma+1)].
$$

$V(\lambda|F_{12})$ is monotonically increasing in λ over $[0,0.5]$, at which point it jumps downward to $-\mu_0$ and remains constant. □

Proof of Proposition 4. We proceed by backward induction, and first analyze what happens in stage 2 given Receiver's belief about model λ. Suppose agents anticipate that the most informative equilibrium (the one derived in Section 2) will be played in stage 2.[13] Principal's expected payoff when the true model is F and Receiver has worldview λ is then $V_P(\lambda|F)$. We show that all equilibria in the first stage are payoff-equivalent to a babbling equilibrium, in that they yield a payoff of $V_P(\lambda_0|F)$ to Principal.

In any informative equilibrium, Principal must send (at least) one message n_{12} that leads Receiver to update her belief downwards—$\lambda(n_{12})<\lambda_0$—and another n_1 that leads Receiver to update her belief upward—$\lambda(n_1)>\lambda_0$. First consider the simpler case in which Receiver's posterior belief is always above λ^*: $\lambda(n)>\lambda^*$ for all $n\in\operatorname{supp} q_P$. Then, irrespective of what message P sends in the first stage, communication with SS in the second stage always yields the same outcome: SS fully reveals the state, and R sets $a_1=\theta_1$ and $a_2=0$. The exact same outcome would have been achieved had P not communicated any information about the model. Indeed, Receiver would have remained at her prior λ_0, which must be above λ^*,[14] and hence interacted in the exact same way with SS.

[13] If they anticipated the babbling equilibrium to be played, then communication on models would be irrelevant and babbling as well.

[14] It is impossible to only induce posteriors $\lambda(n)$ that all lie strictly above the prior.

Now suppose Receiver's posterior is sometimes below λ^*: there exist $n \in \operatorname{supp} q_P$ such that $\lambda(n) \leq \lambda^*$ and call such message n_{12}. Note that Principal can never send that message when the true model is F_1 as sending whichever message leads Receiver to update her belief upward yields a strictly greater payoff: $q_P(n_{12}|F_1) = 0$. But then $\lambda(n_{12}) = 0$, that is, upon receiving message n_{12}, Receiver knows for sure that the true model is F_{12}. That however cannot occur in equilibrium as revealing fully that the model is F_{12} yields the lowest possible payoff for P: $V_P(0|F_{12}) < V_P(\lambda|F_{12})$ for all $\lambda > 0$. Hence P would want to deviate and send whichever message leads to a higher posterior.

□

Proof of Proposition 5. To derive the optimal decision rule under full commitment, we can rely on the Revelation Principle: it is without loss to restrict attention to equilibria under which the Sender announces a state $M = \Theta$ and reports truthfully $q_S(m = \theta|\theta) = 1$. The optimal decision rule $\rho : \Theta \longrightarrow \Delta A$ then solves

$$\max_{\rho} -(1-\mu_0)[2\rho(\overline{a}_1, \overline{a}_2|\underline{\theta}) + \rho(\overline{a}_1, \underline{a}_2|\underline{\theta}) + \rho(\underline{a}_1, \overline{a}_2|\underline{\theta})] - \mu_0[2\rho(\underline{a}_1, \underline{a}_2|\overline{\theta}) + \rho(\overline{a}_1, \underline{a}_2|\overline{\theta}) + \rho(\underline{a}_1, \overline{a}_2|\overline{\theta})]$$

$$\text{s.t.} \; -(\rho(\overline{a}_1, \overline{a}_2|\underline{\theta}) + \rho(\overline{a}_1, \underline{a}_2|\underline{\theta})) + \gamma(\rho(\overline{a}_1, \overline{a}_2|\underline{\theta}) + \rho(\underline{a}_1, \overline{a}_2|\underline{\theta})) \geq -(\rho(\overline{a}_1, \overline{a}_2|\overline{\theta}) + \rho(\overline{a}_1, \underline{a}_2|\overline{\theta})) + \gamma(\rho(\overline{a}_1, \overline{a}_2|\overline{\theta}) + \rho(\underline{a}_1, \overline{a}_2|\overline{\theta}))$$

$$\text{and} \; -(\rho(\underline{a}_1, \overline{a}_2|\overline{\theta}) + \rho(\underline{a}_1, \underline{a}_2|\overline{\theta})) + \gamma(\rho(\overline{a}_1, \overline{a}_2|\overline{\theta}) + \rho(\underline{a}_1, \overline{a}_2|\overline{\theta})) \geq -(\rho(\underline{a}_1, \overline{a}_2|\underline{\theta}) + \rho(\underline{a}_1, \underline{a}_2|\underline{\theta})) + \gamma(\rho(\overline{a}_1, \overline{a}_2|\underline{\theta}) + \rho(\underline{a}_1, \overline{a}_2|\underline{\theta}))$$

First, the probability of taking the high action $a_2 = \overline{a}_2$ must be weakly higher in state $\theta = (1,1)$ than in state $\theta = (0,0)$ under an optimal decision rule, so only the first incentive compatibility constraint binds.

Second, setting $\rho(\underline{a}_1, \overline{a}_2|\overline{\theta}) > 0$ cannot be optimal: any positive weight on action $a = (0,1)$ in the high state can be shifted to action $a = (1,1)$. This relaxes the incentive constraint and strictly increases the objective of Principal. The same is true for for $\rho(\underline{a}_1, \underline{a}_2|\overline{\theta})$: any positive weight on action $a = (0,0)$ in the high state can be shifted to action $a = (1,0)$, yielding a strict improvement. A similar logic yields that, optimally, $\rho(\overline{a}_1, \overline{a}_2|\underline{\theta}) = 0$, $\rho(\overline{a}_1, \underline{a}_2|\underline{\theta}) = 0$, and $\rho(\underline{a}_1, \overline{a}_2|\underline{\theta}) = 0$. Hence, under an optimal decision rule, Receiver chooses action $a = (0, 0)$ with probability one in state $(0, 0)$.

The problem rewrites as

$$\max_{\rho} -\mu_0\rho(\overline{a}_1, \underline{a}_2|\overline{\theta}) \quad \text{s.t. } 0 \geq -\rho(\overline{a}_1, \underline{a}_2|\overline{\theta}) + (\gamma - 1)\rho(\overline{a}_1, \overline{a}_2|\overline{\theta}).$$

Since $\rho(\overline{a}_1, \underline{a}_2|\overline{\theta}) + \rho(\overline{a}_1, \overline{a}_2|\overline{\theta}) = 1$ this yields

$$\rho(\overline{a}_1, \underline{a}_2|\overline{\theta}) = \frac{\gamma - 1}{\gamma}, \text{ and } \rho(\overline{a}_1, \overline{a}_2|\overline{\theta}) = \frac{1}{\gamma}.$$

The very same outcome is achieved without commitment, in the equilibrium derived in Section 2, for $\lambda = \lambda^*$. □

Proof of Proposition 1′. Receiver's optimal action given $\lambda \in \Delta\mathcal{F}$ and μ is $\sigma_k^*(\mu, \lambda) = \mathbb{1}\left\{\mu \geq \frac{1}{2(1-\lambda(k \in F))}\right\}$. Note that if Receiver's worldview puts sufficient weight on models that exclude variable k, i.e., $\lambda(k \in F) < 0.5$, then she never takes the high action: $\sigma_k^*(\mu, \lambda) = 0$ for all μ.

Case 1: $\lambda(1 \in F) < 0.5$ such that $\sigma_1^*(\mu, \lambda) = 0$ for all μ. Since Sender cannot impact Receiver's action on the first dimension and Sender strictly prefers higher action on the second dimension irrespective of the realized state, the only equilibrium is (payoff-equivalent to) babbling.

Case 2: $\lambda(2 \in F) > \lambda(1 \in F) \geq 0.5$. As argued above, Receiver cannot take action $a_2 = 1$ for sure following some message in equilibrium. Hence Receiver's equilibrium posterior must always lie weakly below $0.5[1 - \lambda(2 \in F)]^{-1} < 0.5[1 - \lambda(1 \in F)]^{-1}$. This implies Receiver never takes action $a_1 = 1$ on the equilibrium path, and that communication is effectively only on the second dimension: the only equilibrium is (payoff-equivalent to) babbling.

Case 3: $\lambda(1 \in F) \geq \lambda(2 \in F) \geq 0.5$. Following a similar construction as in the body of the paper, there now exists an informative equilibrium in which two messages are sent, inducing posteriors $\mu(\underline{m}) = 0$ and $\mu(\overline{m}) = 0.5[1 - \lambda(2 \in F)]^{-1}$. The equilibrium is independent of $\lambda(1 \in F)$, and strictly more informative the smaller $\lambda(2 \in F)$.

□

Proof of Proposition 3′. Any worldview that falls under case 1 or case 2 above leads to a babbling equilibrium. Any worldview that falls under case 3 leads to the same equilibrium as in our main analysis. An optimal worldview must then have $\lambda(2 \in F) = 0.5$. It can have any $\lambda(1 \in F) \geq 0.5$, and in particular can have $\lambda(1 \in F) = 1$. This is equivalent to putting weight 0.5 on model F_1 and the rest on model F_{12}.

□

References

Ball, I. (2019). Scoring strategic agents. *arXiv preprint arXiv:1909.01888*.

Bischof, D. and Senninger, R. (2018). Simple Politics for the People? Complexity in Campaign Messages and Political Knowledge. *European Journal of Political Research*, 57(2):473–495.

Chakraborty, A. and Harbaugh, R. (2010). Persuasion by Cheap Talk. *American Economic Review*, 100(5):2361–82.

Che, Y.-K. and Kartik, N. (2009). Opinions as Incentives. *Journal of Political Economy*, 117(5):815–860.

Crawford, V. P. and Sobel, J. (1982). Strategic Information Transmission. *Econometrica*, pages 1431–1451.

Eliaz, K. and Spiegler, R. (2020). A Model of Competing Narratives. *American Economic Review*, 110(12):3786–3816.

Frankel, A. and Kartik, N. (2022). Improving information from manipulable data. *Journal of the European Economic Association*, 20(1):79–115.

Guriev, S. and Papaioannou, E. (2022). The Political Economy of Populism. *Journal of Economic Literature*.

Levy, G. and Razin, R. (2007). On the Limits of Communication in Multidimensional Cheap Talk: A Comment. *Econometrica*, 75(3):885–893.

Levy, G., Razin, R., and Young, A. (2022). Misspecified Politics and the Recurrence of Populism. *American Economic Review*, 112(3):928–62.

Lipnowski, E. and Ravid, D. (2020). Cheap Talk with Transparent Motives. *Econometrica*, 88(4):1631–1660.

Olea, J. L. M., Ortoleva, P., Pai, M. M., and Prat, A. (2022). Competing models. *Quarterly Journal of Economics*.

Schwartzstein, J. and Sunderam, A. (2021). Using Models to Persuade. *American Economic Review*, 111(1):276–323.

Whitmeyer, M. (2019). Bayesian elicitation. *arXiv preprint arXiv:1902.00976*.