American Political Science Review (2020) 114, 3, 724–743
doi:10.1017/S0003055420000209

I Don't Know

MATTHEW BACKUS *Columbia University*
ANDREW T. LITTLE *University of California, Berkeley*

Political decision makers make choices in a complex and uncertain world, where even the most qualified experts may not know what policies will succeed. Worse, if these experts care about their reputation for competence, they may be averse to admitting what they don't know. We model the strategic communication of uncertainty, allowing for the salient reality that sometimes the effects of proposed policies are impossible to know. Our model highlights the challenge of getting experts to admit uncertainty, even when it is possible to check predictive success. Moreover, we identify a novel solution: checking features of the question that only good experts will infer—in particular, whether the effect of policies is knowable—can induce uninformed experts do say "I Don't Know."

> *"[I]t is in the admission of ignorance and the admission of uncertainty that there is a hope for the continuous motion of human beings in some direction that doesn't get confined, permanently blocked, as it has so many times before in various periods in the history of man."*
> *—Richard Feynman, John Danz Lecture, 1963*

> *"Policy-making is hard."*
> *—Callander (2011)*

INTRODUCTION

Political decision makers frequently make disastrous choices. They waste blood and treasure on unwinnable wars, destroy economies with poor monetary policy, and underestimate the threat of coups or revolutions before their opponents show up at the gates. Sometimes poor decisions are made despite the availability of information about how they will turn out. Decision makers may not consult the right experts, or they may ignore their advice. At other times, the best course of action isn't even knowable, and the real danger is being persuaded to take risky action by "experts" who pretend to know what policies will work.

Most work on strategic communication in political science focuses on problems driven by differences of preference ("bias") among experts and decision makers (e.g., Gailmard and Patty 2012; Gilligan and Krehbiel 1987; Patty 2009). However, even if experts have the same *policy* preferences as decision makers do, some are more competent than others are at assessing the available evidence required to give good advice. And, as a literature in economics and finance on career concerns following Holmström (1999) makes clear, experts' desire to appear competent ("reputational concerns") can distort the actions of agents, including the advice they give to principals (e.g., Ottaviani and Sørensen 2006).

We bring this style of communication model to a political context and also place a novel focus on the *difficulty* of policy questions. As is familiar to anyone who has tried to study the causal effect of policies (empirically or theoretically), some questions are harder to answer than others. Uncertainty may be driven by expert incompetence or by the difficulty of the policy question. However, as long as knowledge about the effects of policies is correlated with competence, uninformed experts (competent or not) risk a reputational hit for admitting uncertainty. As a result, experts who care about perceptions of their competence will be reluctant to say "I don't know."

Can this problem be solved by validating experts' claims, by asking other experts, checking other sources, or waiting to see if their predictions come true? Our core contention is that the answer depends on what exactly gets validated. Perhaps the most intuitive kind of validation is what we call *state validation*, or checking whether the expert claims are correct. We find that—at least by itself—this is not effective at getting uninformed experts to report their ignorance. On the other hand, *difficulty validation*, which means checking whether the question was answerable in the first place, tends to be much more effective at inducing experts to admit uncertainty.

We develop a formal model that highlights this problem and our proposed solution in a clear fashion. A decision maker (DM, henceforth "she") consults an expert (henceforth "he") before making a policy decision. The DM is uncertain about a state of the world that dictates which policy choice is ideal. The DM is also

Matthew Backus, Philip H. Geier Jr Associate Professor, Department of Economics, Columbia University, NBER, and CEPR, matthew.backus@columbia.edu

Andrew T. Little, Assistant Professor, Department of Political Science, University of California, Berkeley, andrew.little@berkeley.edu

Thanks to Charles Angelucci, Jonathan Bendor, Sylvain Chassang, Wouter Dessein, Sean Gailmard, James Hollyer, Ryan Hübert, Navin Kartik, Greg Martin, Mallesh Pai, Matthew Mitchell Andrea Prat, Michael Raith, Daniel Rappaport, Maher Said, Jim Snyder, Joel Sobel, Philipp Strack, and audiences at MPSA 2015, EARIE 2017, The 28th International Game Theory Conference, QPEC 2017, Petralia Workshop 2017, SAET 2017, ESSET 2018, Columbia, Harvard, the Higher School of Economics, the University of Hong Kong, Peking University, and Stanford for thoughtful comments and suggestions. We thank Jett Pettus, Alphonse Simon, and Brenden Eum for their excellent research assistance. All remaining errors are our own.

Received: March 13, 2019; revised: November 16, 2019; accepted: April 03, 2020.

uncertain about the quality of the expert and whether the optimal policy is knowable. Experts are either competent (“good”) or incompetent (“bad”), and the question of which policy is better is either answerable (“easy”) or not (“hard”). Bad experts learn nothing about the state of the world. Good experts learn the state if and only if the question is answerable. This means there are two kinds of uninformed experts: bad ones who never learn truth and good ones faced with an unanswerable question.

The expert then communicates a message to the decision maker, who chooses a policy. Finally, the decision maker—potentially endowed, ex post, with information about the state or question difficulty—forms posterior beliefs about the expert quality. The best decisions are made in an *honest* equilibrium, where experts who know which policy is better reveal this information and the uninformed types all say “I don’t know.”

Our main analysis studies the scenario where an expert primarily cares about his reputation. That is, the expert has relatively little concern for the quality of the policy made, though he is unbiased in the sense that the expert and DM agree on which policy is best (conditional on the state). If the DM gets no ex post information about the state or difficulty—the *no validation* case—our answer is bleak. Since they face no chance of being caught guessing, uninformed types could claim to know whether the policy would succeed and appear competent. As a result, honesty is impossible.

What if the DM learns the truth about the ideal policy (state validation) before evaluating the expert? One might expect that state validation can induce honesty, since it is possible to “catch” uninformed experts guessing incorrectly. But what should the DM infer when seeing an incorrect guess: that the expert is incompetent or just that he is uninformed? Under a restriction to strategies and beliefs related to the Markov refinement common to repeated games, we show that the competent uninformed experts and the incompetent uninformed experts must play the same strategy. Further, upon observing an incorrect guess, the DM should infer that the expert is uninformed but still possibly competent (i.e., the exact same inference if the expert said “I don’t know”). Since the expert might get away with guessing and being caught is no worse than admitting uncertainty, an honest equilibrium is still not possible.

The limits of state validation echo past pessimistic results about how reputational concerns limit communication (e.g., Ottaviani and Sørensen 2006). However, our focus on the importance of problem difficulty also suggests a novel path forward. The key barrier to honest communication with state validation is that it does not allow the competent experts asked an unanswerable policy question to differentiate themselves from the incompetent experts. Consider this from the perspective of a competent but uninformed expert: he knows that he is uninformed because it is impossible to know which policy is better, but precisely for this reason he can’t do a better job of guessing the state than an incompetent expert. Where the competent expert does have a definitive advantage over the incompetent expert is not in knowing which policy is better but in knowing whether the ideal policy is knowable in the first place.

We build on this insight to reach our key positive result: if the DM learns ex post whether the question was answerable (difficulty validation), partial if not complete admission of uncertainty is possible.[1] The good uninformed experts admit uncertainty, confident that the DM will learn the ideal policy wasn’t knowable. Bad experts may admit uncertainty as well if this is safer than guessing and potentially getting caught making a prediction when the validation reveals that the question was unanswerable.

These results have practical implications for how political decision makers should structure their interactions with experts. Consulting experts with an interest in good policy being made or investing in methods to check whether their predictions are correct (e.g., running pilot studies) is useful for some purposes, but not for eliciting admission of uncertainty. Rather, it is important for decision makers to be able to eventually learn whether the questions they ask are answerable. We discuss several ways this can be accomplished, such as consulting multiple experts or ensuring decision makers (or someone who evaluates experts) have some general expertise in research methods in order to be able to evaluate what claims are credible.

In addition to highlighting the importance of difficulty validation, we derive several comparative static results. First, incentives to guess are weaker when experts are generally competent. This implies that admission of uncertainty can be more frequent in environments with more qualified experts. Second, when questions are ex ante likely to be answerable, experts face a stronger incentive to guess. So, the quality of policies can be *lower* in environments where the ideal policy is more frequently knowable because this is precisely when bad experts guess the most, diluting the informative value of any message.

RELATED WORK

Uncertainty about ideal policy has at least two causes. First, different people want different things. Even if the effects of policy choices are well known, it may be hard to determine what is best collectively. Second, consensus about the effects of different policies is rare. The world is complicated, and frequently the most credible research gives limited if any guidance about the effects of political decisions.

Decision makers can try to learn about what policies will work in several ways. They can hold debates about policy (Austen-Smith 1990), try to learn from the experience of other polities, or experiment with new policies on a tentative basis (Callander 2011). Either as a part of these processes or separately, they can consult experts employed by the government (staffers, bureaucrats, other politicians) or elsewhere (think tank

[1] As elaborated in the formal analysis, this also requires either non-zero policy concerns or state validation.

https://doi.org/10.1017/S0003055420000209 Published online by Cambridge University Press

TABLE 1. Classification of Related Literature

	Preference bias	Reputation for competence
Expert/advisor	Crawford and Sobel (1982), Gilligan and Krehbiel (1987), Gailmard and Patty (2013)	Ottaviani and Sørensen (2006), Rappaport (2015), *This Paper*
Decision maker	Fearon (1999), Fox and Shotts (2009)	Holmström (1999), Ashworth (2012), Canes–Wrone et al. (2001)

Note: Here we classify related literature by whether the informed party/sender is an advisor (top row) or the actual decision maker (bottom row) and whether the main conflict of interest is different preferences over ideal policy (left column) or reputation for competence (right column).

employees, academics, pundits; Calvert 1985)[2] or delegate to them where optimal (Dessein 2002).[3]

Even if there are experts who have a solid grasp on the wisdom of proposed policies, there are always less competent "experts" who lack valuable information but may still pretend to be informed. And with heterogenous preferences, even good experts may disagree with decision makers about how to best use this information. As a result, the challenges to knowing the ideal course of action in the first place (ignorance and preference bias) generate parallel problems of getting this information into the hands of actual policy makers. Further, if the policy makers themselves have more information than voters and heterogenous competence, the signaling implications of their choices may also lead to suboptimal policies.

The main literatures that we draw on study these problems, and Table 1 presents a classification of the most related work to clarify where our model fits in. The rows correspond to the identity of the "sender," and the columns to whether the main conflict of interest is differing policy preferences or the sender's desire to appear competent.

Most of the political science literature on decision making under uncertainty highlights the problems that arise when actors (experts, policy makers, bureaucrats, voters) have different preferences or ideologies. The top left cell of Table 1 contains examples where, following Crawford and Sobel (1982), an informed advisor or expert (not the decision maker) may not communicate honestly because of a difference in ideal policy. Much of this work has studied how different institutional features like committee rules in congress (Gilligan and Krehbiel 1987), bureaucratic hierarchies (Gailmard and Patty 2012), alternative sources of information (Gailmard and Patty 2013), and voting rules (Schnakenberg 2015) either solve or exacerbate communication problems. Formal models of communication in other political settings such as campaigns (Banks 1990), lobbying (Schnakenberg 2017), and international negotiations (Kydd 2003) also focus on how different preferences affect equilibrium communication.

Our main innovation with respect to the formal theories of expert communication in political science is to bring focus to the other main barrier to communication: reputation concerns (right column of Table 1). Most related work on reputation for competence focuses not on experts but politicians themselves (bottom right cell of Table 1). In some of these models (following Holmström 1999), politicians exert effort (e.g., Ashworth 2012) or otherwise manipulate the information environment (e.g., Little 2017) to make signals of their ability or performance more favorable. Closer to our setting, others model the competence of decision makers as affecting their ability to discern the best policy. In these models, concern about reputation can lead to suboptimal policy choices if, for example, voters think certain policies are ex ante more likely to be favored by competent politicians (Cane–Wrone et al. 2001). The bottom left cell of Table 1 contains examples of models where politicians also want to develop a reputation for being ideologically aligned with citizens (Fearon 1999), which sometimes creates trade-offs with concerns for competence (Fox and Shotts 2009).[4]

We argue that studying the reputation concerns of experts (top right cell of Table 1) is fundamental in political settings. By definition, experts typically have more policy-relevant information than politicians do. Further, particularly in a cheap-talk environment where experts have no direct control over the policy being made (and frequently have a relatively small influence on what choice is made), they sometimes, if not usually, care more about their reputation for competence than the effect of their advice on policy. To our knowledge, there are no other models of "reputational cheap talk" in political science. Related work in other disciplines has shown that reputation concerns lead experts to bias and overstate their reports in order to convince a decision maker that they are the "good" type (e.g., Ottaviani and Sørensen 2006), though the exact kind of lie this induces may depend on the career

[2] See also Manski (2019) for a broader discussions about communication of scientific uncertainty in policy analysis.

[3] See Fehrler and Janas (2019) for a model and experiment on delegation with a focus on competence.

[4] Not all work in the bottom row involves choices made by politicians. For example, Leaver (2009) suggest that bureaucrats with reputation concerns (and a desire to avoid public criticism) may make suboptimal choices (bottom right cell), and judges fear having decisions overturned by higher courts who may have different preferences (e.g., Hübert 2019) (bottom right).

stage of the agent (Prendergast and Stole 1996) or the precise information structure (Rappaport 2015).

In addition to bringing this style of model to a political context, we contribute to the reputational cheap-talk literature by focusing on heterogeneity in the difficulty of questions and introducing the notion of ex post validation of question difficulty rather than whether experts' claims were "correct." In one sense, this is related to the precision of experts' private information in models such as Ottaviani and Sørensen (2006) or their accuracy in related and recent work on screening forecasters by Deb, Pai, and Said (2018)—it creates variation in the quality of signals. However, there is an important feature that differentiates difficulty as we have formulated it: it is a property of the problem itself, not the expert or the expert's signal. This drives our results. Validating the difficulty of problems generates the key informational wedge between good uninformed experts (who know the problem difficulty) and bad experts (who do not).[5]

THE MODEL

Our model is motivated by the following scenario: a decision maker (abbreviated DM, pronoun "she") is making a policy choice. The DM could be the chief executive of a country, a local executive (governor, mayor), or the head of a bureaucracy. There is an unknown state of the world that affects the optimal policy. However, the DM does not observe this state; to this end she employs an expert (pronoun "he"). Experts may be competent or incompetent. Competent experts sometimes know the state of the world, but at other times the state is unknowable. Incompetent experts know nothing.

For concreteness, we will use a running example where the policy in question is how much to restrict the emissions of a chemical, and the DM is the head of an environmental agency with statutory authority to set this regulation level. The expert could be a scientist within the agency or hired as an external consultant. A natural source of uncertainty is whether the chemical in question is harmful to humans. If it is harmful, the DM will want to choose more stringent regulations; if it is not harmful there is no reason to regulate emissions. The expert will learn whether the chemical is harmful if two things are true: (1) he is competent enough to locate and digest the relevant scientific literature and (2) this literature contains an answer to whether the chemical is harmful. If the expert is not competent or there is no scientific consensus on the effect of the chemical, the expert will not learn anything useful.

The Information Environment

We formalize this information structure as follows.

State of the World. Let the state of the world be $\omega \in \Omega \equiv \{0,1\}$. The state of the world encodes the decision-relevant information for the DM (e.g., $\omega = 1$ if the chemical in question is harmful to humans and $\omega = 0$ if not). It is unknown to the DM at the outset, which is why she consults an expert.

Let p_1 represent the common knowledge probability that the state is 1, so it is equal to 0 with probability $1 - p_1$. To reduce the cases to consider, we also assume that $\omega = 1$ is the ex ante more likely state, so $p_1 \geq 1/2$.[6]

Expert Types. The expert has a type $\theta \in \Theta \equiv \{g,b\}$, which indicates whether he is *g*ood (able to digest the relevant scientific literature) or *b*ad (not able to digest the scientific literature). For linguistic variety, we often call good experts "competent" and bad experts "incompetent". Let $p_g \in (0,1)$ be the probability than an expert is good, and so $1 - p_g$ represents the probability of a bad expert. This probability is common knowledge, and the expert knows his type.

Question Difficulty. The difficulty of the question is captured by another random variable, $\delta \in \Delta \equiv \{e,h\}$. That is, the question may be *e*asy (the scientific literature is informative about whether the chemical is harmful), or *h*ard, (the best research does not indicate whether the chemical is harmful). Let p_e be the common knowledge probability that $\delta = e$, so $\delta = h$ with probability $1 - p_e$. The difficulty of the question is not directly revealed to either actor at the outset.

Expert Signal. The expert's type and the question difficulty determine what he learns about the state of the world. In our main analysis, we assume that the expert receives a completely informative signal if and only if he is good *and* the question is easy. If not, he learns nothing about the state. (In section 3 of the Supplemental Information, we analyze a more general information structure, which only assumes that a signal is more likely to be informative when the expert is good and the question is easy). Formally, let the signal be

$$s = \begin{cases} s_1 & \omega = 1, \theta = g \text{ and } \delta = e \\ s_0 & \omega = 0, \theta = g \text{ and } \delta = e \\ s_\varnothing & \text{otherwise.} \end{cases} \qquad (1)$$

In what follows, we will often refer to an expert who observes s_0 or s_1 as *informed* and an expert who observes $s_\varnothing$ as *uninformed.* Importantly, this distinction is not the same as good and bad. If an expert is informed he must be good because bad experts always observe $s_\varnothing$. However, an uninformed expert may be good (if $\delta = h$) or bad.

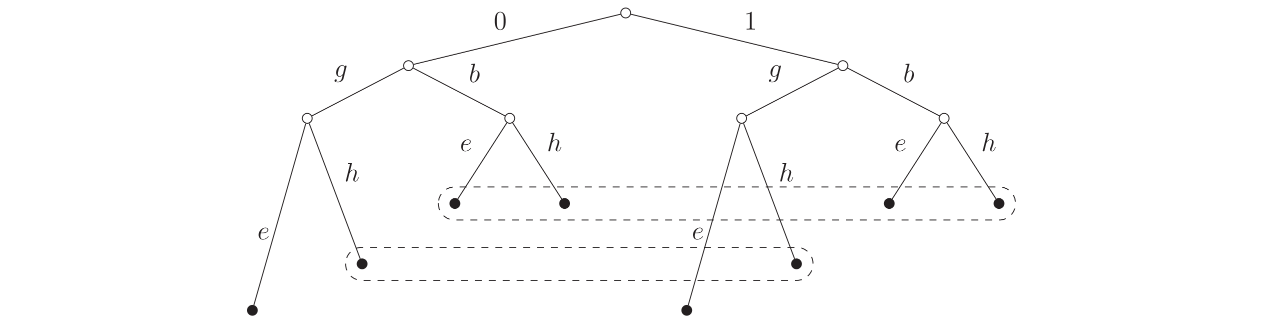
An important implication of our signal structure is that the good expert infers δ from his signal because he observes s_0 or s_1 when $\delta = e$ and $s_\varnothing$ if $\delta = h$. More concretely, a competent expert can review the relevant literature and always determine whether the harmfulness of a chemical is known. On the other hand, bad

[5] The closest to what we are calling difficulty in the prior literature that we are aware of is the information endowment of managers in Dye (1985) and Jung and Kwon (1988), in an altogether different setting where shareholders are uncertain as to the informational endowment of managers.

[6] By the symmetry of the payoffs introduced below, substantively identical results hold if state 0 is more likely.

https://doi.org/10.1017/S0003055420000209 Published online by Cambridge University Press

FIGURE 1. Nature's Play and Experts' Information Sets



experts, who always observe $s_{\varnothing}$, learn nothing about the question difficulty.

Sequence of Play and Payoffs

The game proceeds in the following sequence: first, Nature picks the random variables (ω, θ, δ), according to independent binary draws with the probabilities p_1, p_g, and p_e, specified above. Second, the expert observes his competence and signal and chooses a message from a infinite message space $\mathcal{M}$. The information sets of the expert are summarized in Figure 1. There are four: first, the expert may be bad; second, the expert may be good and the question hard; third, the expert may be good, the question easy, and the state 0; and finally, the expert may be good, the question easy, and the state 1.

Next, the DM observes m and takes an action $a \in [0,1]$, the *policy* choice. Her information set in this stage consists only of the expert report; that is, $\mathcal{I}_{DM1} = (m)$.

In the running example, we can interpret a as the stringency of regulations applied to the chemical in question (restrictions on emissions, taxes, resources to spend on enforcement). To map cleanly to the formalization, $a = 0$ corresponds to the optimal policy if the chemical is not harmful. Policy $a = 1$ corresponds to the optimal level of regulation if the chemical is harmful.

Formally, let $v(a,\omega)$ be the *value* of choosing policy a in state ω. We assume the policy value is given by $v(\mathrm{a}, \omega) \equiv 1 - (a - \omega)^2$. If taking a decisive action of $a = 0$ or $a = 1$, the value of the policy is equal to 1 for making the correct choice ($a = \omega$) and 0 for making the wrong choice ($a = 1 - \omega$). Taking an interior action ($0 < a < 1$) gives an intermediate policy value, where the v function implies an increasing marginal cost the further the action is from the true state. This function is common knowledge, but since the DM may be uncertain about ω, he may be uncertain about how the policy will turn out and hence the ideal policy. Let $\pi_1 = \mathbb{P}(\omega = 1|\mathcal{I}_{DM1})$ denote the DM's belief that $\omega = 1$ when he sets the policy. Then, the expected value of taking action a is

$$1 - \left[\pi_1(1 - a)^2 + (1 - \pi_1)a^2\right], \tag{2}$$

which is maximized at $a = \pi_1$.

The quadratic loss formulation conveniently captures the realistic notion that when the expert does not learn the state, the decision maker makes a better policy choice (on average) when learning this rather than being misled into thinking the state is zero or one. That is, in our formalization it is best to pick an intermediate level of regulation when it is unclear whether the chemical is harmful (e.g., modest emissions restrictions, labeling requirements). Formally, if the question is unsolvable, the optimal action is $a = p_1$, giving an average payoff of $1 - p_1(1 - p_1)$, which is strictly higher than the average value of the policy for any other action.

After the policy choice is made, the DM may receive some additional information (validation), and she then forms an inference about the expert competence $\pi_g \equiv \mathbb{P}(\theta = g|\mathcal{I}_{DM2})$. Different validation regimes (described below) affect what information the DM has at this stage, $\mathcal{I}_{DM2}$.

The decision maker only cares about the quality of the policy:

$$u_{DM} = v(\mathrm{a}, \omega). \tag{3}$$

So, as derived above, the DM will always pick a policy equal to her posterior belief that the state is 1, π_1.

The expert cares about the belief that he is competent (π_g) and potentially also about the quality of the policy choice. We parameterize his degree of *policy concerns* by $\gamma \geq 0$ and write his payoff

$$u_E = \pi_{\mathrm{g}} + \gamma v(a,\omega). \tag{4}$$

We first consider the case where $\gamma = 0$; that is, the expert only cares about his reputation. We then analyze the case where $\gamma > 0$, focusing attention on the case where policy concerns are small ($\gamma \to 0$) in the main text.

Validation

Finally, we formalize how different kinds of ex post validation affect what the DM knows ($\mathcal{I}_{DM2}$) when forming a belief about the expert competence. For our regulation example, there are several ways the DM might later learn things about the state or difficulty of the question.

https://doi.org/10.1017/S0003055420000209 Published online by Cambridge University Press

First, there may be studies in progress which will later clearly indicate whether the chemical is harmful. Alternatively, if the question is whether the chemical has medium or long-term health consequences, this may not become clear until after the initial regulation decisions are made. To pick a prominent contemporary example, the degree to which we should regulate "vaping" of nicotine and cannabis depends on the long-term health consequences of this relatively new technology —perhaps relative to smoking—which are not currently known.

Second, the DM could consult other experts (who may themselves have strong or weak career concerns). These other experts could be asked about the state of the world or perhaps the quality of the evidence on dimensions the DM may not be able to assess (Anecdotal evidence or scientific? Randomized control trials or observational studies? Have they been replicated? Human or animal subjects?).

In other contexts, validation about the state of the world could naturally arise if it is about an event that will later be realized (the winner of an election, the direction of a stock's movement) but where the DM must make a choice before this realization. Alternatively, in many policy or business settings, the decision maker may be able to validate the expert's message directly, whether through experimental A/B testing or observational program evaluation. Closer to difficulty validation is the familiar notion of "peer review," whereby other experts evaluate the feasibility of an expert's design without attempting the question themselves. Another possibility is when the decision maker has expertise (substantive or methodological) in the general domain of the question but does not have the time to investigate herself, so she may be able to validate whether the question was answerable only after seeing what the expert comes up with.

Alternatively, subsequent events may reveal auxiliary information about whether the state should have been knowable, such as an extremely close election swayed by factors which should not have been ex ante predictable (i.e., this reveals that forecasting the winner of the election was a difficult question).

Formally, we consider several variations on $\mathcal{I}_{DM2}$. In all cases, the structure of $\mathcal{I}_{DM2}$ is common knowledge.

In the *no validation* case, $\mathcal{I}_{DM2} = (m)$. This is meant to reflect scenarios where it is difficult or impossible to know the counterfactual outcome had the decision maker acted differently. The *state validation* case, $\mathcal{I}_{DM2} = (m, \omega)$, reflects a scenario in which the DM can check the expert's advice against the true state of the world. In the *difficulty validation* case, $\mathcal{I}_{DM2} = (m, \delta)$, meaning that the DM learns whether the question was hard (i.e., whether the answer could have been learned by a good expert). In the *full validation* case, $\mathcal{I}_{DM2} = (\text{m}, \omega, \delta)$, the DM learns both the state and the difficulty of the question.

To be clear, many of the motivating examples blend validation about the state and difficulty. We consider the cases of "pure" difficulty or state validation to highlight what aspects of checking expert claims are most important for getting them to admit uncertainty.

EQUILIBRIUM DEFINITION AND PROPERTIES

The standard solution concept for a model like ours (with sequential moves and incomplete information) is perfect Bayesian equilibrium (PBE). While our central conclusions hold when using this solution concept, they become much clearer when we add a refinement that builds on the Markov restriction common to dynamic games of complete information (Maskin and Tirole 2001). In this section, we define the equilibrium concept in the context of our game; Appendix A contains a more general definition and detailed discussion of how the results change when using PBE. Appendix A also contains a discussion of past usage of related refinements; in short, there are several applied papers that use the same restriction of strategies that we employ, but to our knowledge this is the first paper to make use of the refinement to beliefs that the Markov strategies restriction implies.

Markov Sequential Equilibrium (MSE)

The Markov restriction in repeated games requires that if two histories of play (h_1 and h_2) result in a strategically equivalent scenario starting at both histories (meaning that the players have the same expected utilities over the actions taken starting at both h_1 and h_2), then the players must use the same strategies starting in both histories. The analogous restriction in our setting has implications for strategies and beliefs.

In terms of strategies, we require that if two *types* of expert face a strategically equivalent scenario, they must play the same strategy. Most important to our model, in some cases a competent but uninformed expert and an incompetent expert have the exact same expected utility (for any DM strategy). Perfect Bayesian equilibrium would allow these two types to play different strategies, despite the fact that they face identical incentives; the Markov restriction requires them to play the same strategy. Our restriction to beliefs essentially assumes that, even when observing an off-path message, the DM still believes that the expert plays some Markov strategy.

Formally, let $\sigma_{\theta,s}(m)$ be the probability of sending message m as a function of the sender type, $\pi_1(m)$ be the posterior belief that the state is 1 given message m, $\pi_g(m; \mathcal{I}_{DM2})$ be the posterior belief about the expert's competence, and $a(m)$ be the policy action given message m. Let U_E and U_{DM} be the expected utilities for each player.

Perfect Bayesian equilibrium requires that both players maximize their utility given the other's strategy and their beliefs and that these beliefs are formed by Bayes' rule when possible. To these we add two requirements:

Definition 1. *A Markov sequential equilibrium to the model is a PBE which also meets the following requirements:*

https://doi.org/10.1017/S0003055420000209 Published online by Cambridge University Press

- *(Expert Markov Strategies): If there are two types* (θ', s') *and* (θ'', s'') *such that* $U_E(m;\theta',s') = U_E(m;\theta'', s'')$ *for all* m *given the DM strategies and beliefs, then* $\sigma_{\theta',s'}(m) = \sigma_{\theta'',s''}(m)$ *for all* m.
- *(DM Markov Consistency): For all* m *and* $\mathcal{I}_{DM2}$, *there exists a sequence of non-degenerate Markov strategies for the expert* σ^k, *with corresponding beliefs formed by Bayes' Rule* $\pi_1^k(m)$ *and* $\pi_g^k(m, \mathcal{I}_{DM2})$ *such that* $\pi_1(m) = \lim_{k\to\infty} \pi_1^k(m)$ *and* $\pi_g(m, \mathcal{I}_{DM2}) = \lim_{k\to\infty} \pi_g^k(m, \mathcal{I}_{DM2})$.

The key implication of Markov consistency is to rule out off-path inferences about payoff-irrelevant information, because off-path beliefs that condition on payoff-irrelevant information cannot be reached by a sequence of Markov strategies. Our restriction is related to that implied by D1 and the intuitive criterion (Cho and Kreps 1987). However, where these refinements require players to make inferences about off-path play in the presence of strict differences of incentives between types, our restriction rules out inference about types in the absence of strict differences of incentives.

We use this solution concept for two reasons. The first is theoretical. People have endless "private information" that they could, in principle, condition their behavior on. In a more complex but realistic version of our model, the expert may have private information about not only what he learns about the chemical in question but also about his personal policy preferences, views on what kinds of scientific evidence are credible, and what he had for breakfast in the morning. When observing an off-path message, PBE would allow for updating (or not) on any of these dimensions regardless of their relevance to the interaction at hand. The Markov strategies restriction provides a principled and precise justification for which kinds of private information experts can condition their strategies on[7] (Beliefs about the effects of the chemical in question? Usually. What evidence is credible? Sometimes. Breakfast? Rarely[8]). The Markov beliefs restriction is then a logical implication of what observers can update on when observing off-path messages.

The second reason is more practical: our main result about the importance of difficulty validation for getting experts to admit uncertainty is nearly immediate when using this restriction. As demonstrated in Appendix A, similar results hold when analyzing PBE, but since the set of PBE is much larger, the comparisons are not as clean.

Properties of Equilibria

Since we allow for a generic message space, there will always be many equilibria to the model even with the Markov restrictions. To organize the discussion, we will focus on how much information can be conveyed (about ω and θ). On one extreme, we have babbling equilibria, in which all types employ the same strategy, and the DM learns nothing.

On the other extreme, there is never equilibrium with full separation of types. To see why, suppose there is a message that is *only* sent by the good but uninformed types $m_{g,\varnothing}$ ("I don't know because the optimal policy isn't clear") and a different message only sent by the bad uninformed types $m_{b,\varnothing}$ ("I don't know because I am incompetent"). If so, the policy choice upon observing these messages would be the same. However, the reputation payoff for sending $m_{g,\varnothing}$ is strictly higher, so the bad types have an incentive to deviate.

Still, such full separation is not required for the expert to communicate what he knows about the *state*. That is, there may still be equilibria where the uninformed types say "I don't know" (if not why), and the informed types report the state of the world. We call these *honest* equilibria. In the main text, we focus on the simplest version of an honest equilibrium, where there is a unique message m_x sent by each type observing s_x with probability 1, $x \in \{0, 1, \varnothing\}$; see Appendix B for a more general definition. This is a particularly important class of equilibria in our model because it conveys the most information about the state:

Proposition 1. *The expected value of the policy in an honest equilibrium is* $p_g p_e + (1 - p_g p_e)[1 - p_1(1 - p_1)] \equiv \bar{v}$, *which is strictly greater than the expected value of the policy in any equilibrium that is not honest.*

Proof. Unless otherwise noted, all proofs are in Appendix B. □

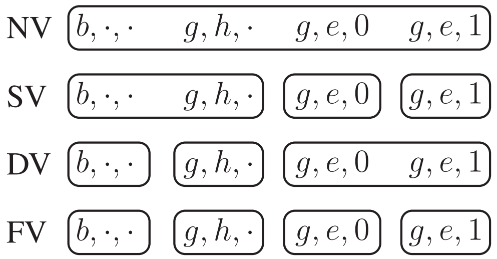
This result formalizes our intuition that it is valuable for the DM to learn when the expert is uninformed, and it follows directly from the convexity of the loss function for bad policies. For example, if the expert on environmental policy sometimes says that a chemical is harmful (or is not harmful) when the truth is that he isn't sure, the regulator will never be entirely sure what the optimal policy is. Her best response to this garbled message is to not regulate as aggressively as she would if knowing the chemical is harmful for sure, even when the expert fully knows that this is the case.

As we will see, honest equilibria tend to fail when one of the uninformed types would prefer to "guess," mimicking the message of the informed types. So, the other equilibria we consider in the main text are ones where the informed types still send their respective messages m_0 and m_1, but the uninformed types at least sometimes send one of these messages. We refer to sending one of these messages as "guessing," and sending $m_\varnothing$ as "admitting uncertainty." See Appendix B for a definition of what it means to admit uncertainty with more general messaging strategies and section 2 of the Supplemental Information for an extensive discussion of why focusing on this class of messaging strategies sacrifices no meaningful generality for our results.

[7] Of course, a classic interpretation of mixed strategies is that the sender conditions on some random (and "irrelevant") information like a coin flip. However, as discussed in Appendix A, if mixed strategies are viewed as the limit of pure strategies with payoff perturbations a la Harsanyi (1973), then only Markov strategies are possible.

[8] Though see Cho and Kreps (1987, section II).

FIGURE 2. Payoff-Equivalence Classes With No Policy Concerns



Note: This figure depicts equivalence classes under each validation regime for the case with no policy concerns. Each row represents a validation regime: respectively, no validation, state validation, difficulty validation, and full validation. Each column represents an expert information set.

Combined with proposition 1, these definitions highlight why admission of uncertainty is important for good decision-making. In any equilibrium with guessing, the fact that the uninformed types send messages associated with informed types leads to worse policies than in an honest equilibrium. This is for the two reasons highlighted in our opening paragraph. First, when the expert actually is informed, his advice will be partially discounted by the fact that the DM knows some uninformed types claim to know which policy is best. That is, decision makers may ignore good advice. Second, when the expert is uninformed, he will induce the DM to take more decisive action than the expert's knowledge warrants; that is, decision makers may *take* bad advice.

WHEN IS ADMISSION OF UNCERTAINTY POSSIBLE?

Between the four possible validation regimes and whether the expert exhibits policy concerns, there are many cases of the model to analyze. In this section we show that the Markov restrictions and some simple incentive compatibility constraints quickly allow us to see when admission of uncertainty is impossible regardless of the particular parameter values. We then provide a complete analysis of one case where admission of uncertainty is possible but not guaranteed for more subtle comparative static results. In other words, we first highlight our negative results about state validation (and discuss why they do not afflict difficulty validation), and we then derive concrete positive results about difficulty validation in the next section. Section 1 of the Supplemental Information contains a full analysis of the remaining cases.

Markov sequential equilibrium has two main implications: Markov strategies, which requires that payoff-irrelevant information cannot affect equilibrium play, and Markov consistency, which requires that off-path beliefs cannot update on payoff-irrelevant information. To see the immediate implications of these restrictions, it is helpful to construct the classes of payoff-equivalent information sets. We put information sets in the same payoff-equivalence class if experts at those decision nodes are payoff equivalent *for any DM strategy*.[9]

Figure 2 illustrates the case with no policy concerns. Each row represents an assumption on the DM's information set at the end of the game, $\mathcal{I}_{DM2}$. Each column represents one of the four information sets depicted in Figure 1.

No Validation, $\gamma = 0$. First, in the no validation (NV) case, $\mathcal{I}_{DM2} = (m)$. Since the DM only sees the message when evaluating the expert, if the expert has no policy concerns his private signal is payoff-irrelevant. Therefore, there is a single payoff-equivalence class comprised of all four information sets, as depicted in the first row of Figure 2, and the Markov strategies restriction implies that all experts play the same strategy. In this case, all that is left is a babbling equilibrium.

Proposition 2. *With no validation and no policy concerns (i.e., $\gamma = 0$), any MSE is babbling, and there is no admission of uncertainty.*

This highlights the importance of the Markov strategies restriction. In our main example, if the environmental expert does not care at all about whether a good policy choice is made (unlikely if he is a citizen who is affected by the policy) *and* there is no check on his claim about whether the chemical is harmful (again, unlikely), the information he has is not payoff relevant. We should not be surprised that this extreme premise leads to an extreme prediction that such an expert will not convey any useful information.

State Validation, $\gamma = 0$. A promising direction is to allow the DM to observe ω when forming beliefs about θ. Doing so breaks payoff equivalence between types with different information about the state, as we illustrate in the second row of Figure 2. This partition of payoff equivalence is rich enough to support honesty in Markov strategies. Bad experts and uninformed good experts can pool on a message interpreted as "I don't know," and informed experts can send messages interpreted as "the optimal policy is zero" and "the optimal policy is one."

To see this, consider the expected payoff for the expert with type and signal (θ, s) sending message m:

$$\sum\nolimits_{\omega \in \{0,1\}} Pr(\omega|s,\theta)\pi_g(m,\omega).$$

Now, the informed types with different information about the state are not generally payoff equivalent since $Pr(\omega|\theta,s)$ depends on the signal. However, good and bad uninformed experts—$(g, s_\varnothing)$ and $(b,s_\varnothing)$—are always payoff equivalent because $Pr(\omega|g, s_\varnothing) = Pr(\omega|b, s_\varnothing)$. So, the Markov *strategies* restriction does not preclude admission of uncertainty (or even an honest equilibrium). However, the Markov consistency restriction does.

To see why, consider the simplest case of an honest equilibrium where types observing signal s_x send message m_x. In such an equilibrium, the decision maker picks a stringent regulation when the expert says the chemical is harmful ($a = 1$ when observing m_1), no regulation when the chemical is not harmful ($a = 0$ when observing m_0), and intermediate regulation when the expert admits not knowing whether the chemical is harmful ($a = p_1$ when observing $m_\varnothing$). The on-path information

[9] Any two information sets can be payoff equivalent for *some* DM strategy: e.g., if she always picks the same policy and competence assessment for all messages.

sets include cases where a good expert makes an accurate recommendation, $(m_0, 0)$ and $(m_1, 1)$, and cases where an uninformed expert (good or bad) says "I don't know" along with either validation result: $(m_\varnothing, 0)$ or $(m_\varnothing, 1)$. When validation indicates the expert was right about whether the chemical is harmful—$(m_0, 0)$ or $(m_1, 1)$—the DM knows the expert is good—$\pi_g(m_i, i) = 1$ for $i \in \{0, 1\}$. When observing $m_\varnothing$ and either validation result, the belief about the expert competence is

$$\begin{aligned}\pi_g(m_\varnothing, \omega) &= \frac{Pr(\theta = g, \delta = h)}{Pr(\theta = g, \delta = h) + Pr(\theta = b)} \\ &= \frac{p_g(1 - p_e)}{p_g(1 - p_e) + 1 - p_g} \\ &\equiv \pi_g^\varnothing.\end{aligned}$$

The term $\pi_g^\varnothing$, which recurs frequently throughout the analysis, represents the share of uninformed types who are competent but asked a hard question.

For the uninformed types, the expected payoff for sending $m_\varnothing$ in the honest equilibrium is $\pi_g^\varnothing$. Since $0 < \pi_g^\varnothing < p_g$, the expert revealing himself as uninformed leads to a lower belief about competence than the prior, but it is not zero because there are always competent but uninformed types.

Consider a deviation to m_1—claiming to know that the chemical is harmful. When validation reveals this to be true, the DM observes $(m_1, 1)$ and believes the expert to be competent with probability 1. When validation reveals the chemical is not harmful, the DM observes $(m_1, 0)$, which is off-path.

However, MSE places some restriction on this belief, as it must be the limit of a sequence of beliefs consistent with a sequence of Markov strategies. Since the good and bad uninformed types are payoff equivalent and play the same strategy, the worst inference that the DM can make about the expert when observing an off-path message/validation combination is that the expert was uninformed; that is, $\pi_g(m_1, 0) \geq \pi_g^\varnothing$ (see the proof of proposition 3). Given this restriction on the off-path belief, in any honest MSE the payoff to sending m_1 must be at least

$$p_1 + (1 - p_1)\pi_g^\varnothing > \pi_g^\varnothing.$$

The expert can look no worse from guessing that the chemical is harmful and being incorrect than he would when just admitting he is uncertain. Since there is a chance to look competent when guessing and being correct, the expert will always do so. This means there is always an incentive to deviate to m_1 (or, by an analogous argument, m_0), and hence no honest equilibrium.

A related argument implies that there is no MSE where the uninformed types *sometimes* admit uncertainty—when $\sigma_\varnothing(m_\varnothing) \in (0, 1)$—and sometimes guess m_0 or m_1: guessing and being correct always gives a higher competence evaluation than incorrectly guessing, which gives the same competence evaluation as sending $m_\varnothing$. So, guessing gives a strictly higher payoff than admitting uncertainty.

Proposition 3. *With state validation and no policy concerns, there is no MSE where an expert admits uncertainty.*

As shown in the proof of the proposition, there is an MSE where the informed types reveal their information. For example, the experts who know the chemical is harmful report this and those who know it is not harmful say so. However, all of the uninformed experts—competent or not—will make a guess at whether the chemical is harmful. The equilibrium condition is that their guessing probabilities are such that observing a claim that the chemical is harmful or not leads to the same belief about the expert competence. So, state validation does improve communication in general relative to no validation but not in terms of admitting uncertainty.

Further, as shown in section 1.2 of the Supplemental Information, adding policy concerns (to the no validation or state validation case) will not solve this problem unless the expert cares so much about the policy as to accept the hit to his reputation from admitting uncertainty.

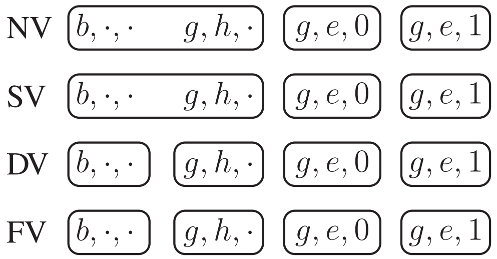
Difficulty, Full Validation, and Policy Concerns. For the DM to effectively threaten punitive off-path beliefs, we need to break the payoff equivalence of bad types and good but uninformed types, and this is precisely what difficulty validation (DV) does, depicted in the third row of Figure 2. However, difficulty validation is not enough to sustain honesty because (unlike state validation) it does not break the payoff equivalence between the informed experts who actually know whether the chemical is harmful.

This we view as a more minor problem, which can be solved by either combining state and difficulty validation (FV), as in the fourth row of Figure 2, or by adding small policy concerns, which yields the payoff-equivalence classes that are represented in Figure 3. We formally prove results about when (complete) communication of uncertainty is possible in the next section.

Summary. From the payoff-equivalence classes alone, we now know what cases at least allow for the possibility of an honest equilibrium without strong policy concerns. First, we need to break the payoff equivalence between the types with different information about whether the chemical is harmful, which can be accomplished with either state validation or any policy concerns. Second, we need to break the payoff equivalence between good and bad uninformed experts, which, given the natural way we have set up the problem, can *only* be accomplished with difficulty validation (or full validation, which includes difficulty validation).

What remains is to check when honesty is in fact possible. In the next section, we do this for the "hardest" case, with only difficulty validation and with small policy concerns. As shown in section 1.2 of the Supplemental Information, the insights from this analysis are similar to what we obtain with larger policy concerns and/or full validation.

FIGURE 3. Payoff-Equivalence Classes With Policy Concerns



Note: This figure depicts equivalence classes under each validation regime for the case with policy concerns. Each row represents a validation regime: no validation, state validation, difficulty validation, and full validation, respectively. Each column represents an expert information set, as derived in Figure 1.

ANALYSIS OF OUR MAIN CASE

While the previous section shows that difficulty validation is necessary for the admission of uncertainty (absent large policy concerns), we have not yet shown when it is sufficient. This section explores when difficulty validation, combined with small policy concerns, is sufficient for honesty, or at least some admission of uncertainty.

We analyze difficulty validation rather than full validation in the main text not because we believe the former is necessarily more common but to show circumstances under which the minimal validation regime is sufficient to induce admission of uncertainty.

The focus on small policy concerns is for both technical and substantive reasons. If the expert has exactly no policy concerns, then this renders the two informed types (i.e., those who know the chemical is harmful or not) payoff equivalent, which greatly undermines the amount of information that can be transmitted in an MSE. However, this is fragile to the small and entirely realistic perturbation where the expert has any nonzero concern about the quality of the policy. For example, even if the chemical in question only has a potentially small effect on the environment, the expert himself could be harmed by exposure. So, having small policy concerns allows the informed types to reveal their information honestly, while simplifying the analysis of the potential deviations for uninformed types since they (in the limit as $\gamma \to 0$) will send the message that maximizes their reputation for competence.

This also hints at the more substantive justification. In many, if not most policy-making domains, the effect of the policy change that experts will feel in their own personal life likely pales in comparison to their concern about perceptions of their competence, which can affect whether they keep their job or will be hired in the future. In general, we expect that our analysis applies more to "big" policy questions that affect many people and where experts are very specialized and care about perceptions of their competence in that particular domain.

Equilibrium

We continue to study the most straightforward class of messaging strategies where the informed types send one message each, to which we give the natural labels m_0 and m_1, and the uninformed types either send one of these messages or a third message labeled $m_\varnothing$ ("I Don't Know").[10]

In an honest equilibrium, messages m_0 and m_1 are only sent by informed types. So, when observing these messages, the DM picks actions $a = 0$ and $a = 1$, respectively. At the evaluation stage, when observing (m_0, e) or (m_1, e), the DM knows the expert is competent with certainty.

Upon observing $m_\varnothing$, the DM picks policy p_1. The competence assessment when the expert says "I don't know" depends on the result of the validation. When the problem is easy, the DM knows that the expert is incompetent because this fact means that a competent expert would have learned whether the chemical is harmful and sent an informative message. So, $\pi_g(m_\varnothing, e) = 0$. Upon observing $(m_\varnothing, h)$, the DM learns nothing about the expert competence because no expert gets an informative signal when the scientific literature is uninformative.

Combining these observations, the payoff to the good but uninformed type for sending $m_\varnothing$ (who knows the validation will reveal $\delta = h$) is

$$p_g + \gamma[1 - p_1(1 - p_1)].$$

The bad uninformed type does not know if the validation will reveal the problem is hard and so receives a lower expected competence evaluation and hence payoff for sending $m_\varnothing$:

$$(1 - p_e)p_g + \gamma[1 - p_1(1 - p_1)].$$

Since no types are payoff equivalent, the Markov consistency requirement places no restrictions on off-path competence evaluations, and we can set these to zero.[11] In the case with only difficulty validation, these are the information sets (m_0, h) and (m_1, h)—getting caught guessing when validation reveals that the scientific literature is not informative. Importantly, the expert is not caught because the DM realizes his claim is wrong but because she comes to learn that the question was not answerable. As formalized below, such an off-path belief is also "reasonable" in the sense that the bad uninformed types face a strictly stronger incentive to guess than the good uninformed types do.

If the DM believes that the expert is bad with probability one upon observing (m_0, h) or (m_1, h), then a good but uninformed type knows he will get a reputation payoff of zero if sending either of these messages. Further, if claiming he knows whether the chemical is

[10] See Section 2 of the Supplemental Information for an extensive discussion of the sense in which this is without loss of generality.

[11] This belief can be reached by assuming a sequence of strategies where only the bad type sends a particular message, but with a probability which approaches zero.

harmful, the policy is worse as well, so he has no incentive to deviate. A bad type guessing that the chemical is harmful (m_1) gets the expected payoff:

$$p_e + \gamma p_{1,}$$

which is strictly higher than the payoff for sending m_0. So, the constraint for an honest equilibrium is

$$(1 - p_e)p_g + \gamma[1 - p_1(1 - p_1)] \geq p_e + \gamma p_1$$
$$\gamma \geq \frac{p_e\left(1 + p_g\right) - p_g}{(1 - p_1)^2}.$$

Not surprisingly, when policy concerns are very strong, uninformed experts will admit uncertainty because it leads to better policy choices. In fact, this holds for any validation regime; see Section 1 of the Supplemental Information for a comparison of "how strong" policy concerns must be to induce honesty for each case. However, this analysis also highlights that when policy concerns are small relative to reputation concerns, there is never an honest equilibrium with no validation or just state validation. In contrast, with just difficulty validation,[12] as $\gamma \rightarrow 0$ there is an honest equilibrium when

$$p_e \leq \frac{p_g}{1 + p_g}. \tag{5}$$

This inequality implies that an honest equilibrium is easy to sustain when p_e is low and p_g is high. The former holds for two reasons: a prior belief that the literature is likely not informative about the chemical means that uninformed experts are likely to be competent, and the uninformed expert is more likely to be "caught" claiming to have an answer to an impossible problem. A situation with more competent experts (high p_g) makes honesty easier because it means the uninformed types are frequently competent, making admitting uncertainty look less bad.

What if Equation 5 does not hold? Of course, there is always a babbling equilibrium, and there is also an always guessing equilibrium where the good and bad uninformed types always send m_0 or m_1. More interesting for our purposes, there can also be an MSE where all of the good types report honestly and the bad types play a mixed strategy over (m_0, m_1, $m_\varnothing$).[13] There is a more subtle incentive compatibility constraint that must be met for this equilibrium to hold: if the bad types are indifferent between sending m_0 and m_1, it can be the case that the *informed* types prefer to deviate to sending the other informed messages (i.e., the s_1 type prefers to send m_0) when policy concerns get very small. See the proof of proposition S.5 in section 1 of the Supplemental Information for details. In short, if the probability of a solvable problem is not too low or the probability of the state being 1 is not too high, then this constraint is not violated and there is an MSE where all of the good types send their honest message.[14]

Proposition 4. *As $\gamma \rightarrow 0$ with difficulty validation, there is an honest MSE if and only if $p_e \leq \frac{p_g}{1+p_g}$. If not, and $p_e \geq 2p_1 - 1$, then there is an MSE where the good types send their honest message and the bad types use the following strategy*:

$$\sigma_b^*(m_\varnothing) = \begin{cases} \dfrac{1 - p_e\left(1 + p_g\right)}{1 - p_g} & p_e \in \left(\dfrac{p_g}{1 + p_g}, \dfrac{1}{1 + p_g}\right) \\ 0 & p_e \geq \dfrac{1}{1 + p_g}, \end{cases},$$
$$\sigma_b^*(m_0) = (1 - p_1)\left[1 - \sigma_b^*(m_\varnothing)\right],$$
$$\sigma_b^*(m_1) = p_1\left[1 - \sigma_b^*(m_\varnothing)\right]. \tag{6}$$

In sum, other than in a restrictive corner of the parameter space,[15] small policy concerns and difficulty validation are sufficient to induce at least good and uninformed experts to admit uncertainty—and potentially full honesty among all experts.

When Do We Observe Admission of Uncertainty, and from Whom?

Figure 4 illustrates the equilibrium described by proposition 4, which helps provide some more concrete insights into when we should expect to see admission of uncertainty and good decisions. All claims that follow are comparisons within this equilibrium, though again this is the only MSE where the good types report honestly (see footnote 13).

In the bottom right corner (most experts are competent, most problems are hard), there is an honest equilibrium. Substantively, this corner of the parameter space plausibly corresponds to environments where the best experts are tackling questions on the research frontier, be it at conferences or in the pages of academic journals. In general, the admission of uncertainty is quite frequent in these settings: even good experts often don't know the answer but feel comfortable admitting this, and as a result bad experts can pool with them and admit uncertainty too. So, while decision makers may not always get useful advice here—recall, most questions have no good answer—they will at least

[12] Unsurprisingly, this constraint is even easier to meet with full validation; the key point is that difficulty validation is *necessary* for honesty with small policy concerns, and it is sometimes sufficient.

[13] This is the only MSE where the good types report honestly, and we conjecture that the equilibrium we check for (when it exists) maximizes both the probability that the expert admits uncertainty and the expected value of the decision. It is hypothetically possible, though we believe unlikely, that there could be an MSE where the good types do not report honestly and this induces the bad types to admit uncertainty more often in a manner that outweighs the loss of information from the good types.

[14] The proof of the proposition discusses how the intuition behind this constraint.

[15] The only time this does not hold is if $\frac{p_g}{1+p_g} < p_e$ and $p_e < 2p_1 - 1$. So, three sufficient conditions for this constraint to remain unviolated are (1) p_g is high, (2) p_1 is high, and (3) p_e is either low or high. Even if both of these inequalities hold, as shown in the proof of proposition S.5 in section 1 of the Supplemental Information, it is a necessary but not sufficient condition for an incentive compatibility constraint to be violated.

have no problems getting experts to honestly report what they know.

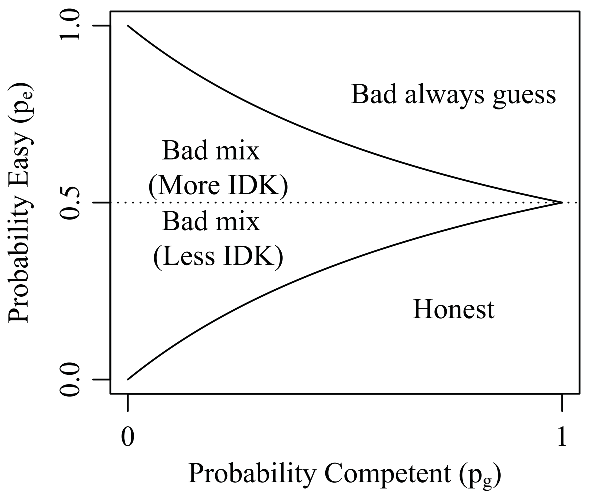
In the top right corner, most experts are competent and most problems are easy. We can think of this part of the parameter space as "business as usual" in bureaucracies, companies, or other organizations where qualified experts are addressing relatively mundane problems that they generally know how to solve. In this case, there is little admission of uncertainty, both because experts typically get informative signals and because admitting uncertainty in this setting is too risky for bad experts to ever do it. While experts have the most information in this region, this comes at a cost from the perspective of the decision maker, as bad types always guess, which dilutes the value of the informative messages.

Comparing across these two cases also hints at the question of when good or bad experts are more likely to admit uncertainty. Our model contains a straightforward assumption about which experts are likely to *be* uncertain about the state of the world: good experts sometimes and bad experts always. And, in an honest equilibrium—when $p_e < p_g/(1+p_g)$—this is what they will report, so bad experts admit uncertainty more often.

However, on the other extreme, when $p_e > 1/(1+p_g)$, it is *only* the good experts who will ever admit uncertainty. Put another way, if an outside observer (who the expert does not care to impress) were to see an expert admit uncertainty in this part of the parameter space (and in this equilibrium), she would know for sure that anyone who says "I don't know" is in fact competent.

More generally, good experts admit uncertainty whenever the problem is hard, with probability $1 - p_e$. For parameters where good experts sometimes admit uncertainty (the triangle in the left of Figure 4), bad experts send $m_\varnothing$ with probability $\frac{1 - p_e(1 + p_g)}{1 - p_g}$, which is less than $1 - p_e$ if and only if $p_e > 1/2$. So, in domains of difficult problems, bad experts are more likely to admit uncertainty, and in domains with easier questions good experts are more likely to admit uncertainty.

FIGURE 4. Illustration of Expert Strategies in the Equilibrium Identified by Proposition 4, as a Function of p_e and p_g, with $p_1 = 1/2$.



Comparative Statics

We now ask how changing the probability parameters of the model affects the communication of uncertainty by uninformed experts in the MSE identified by proposition 4.

In general, one might expect that adding more competent experts will lead to less admission of uncertainty and better decisions, and that making problems easier will also lead to less admission of uncertainty and better decisions. Both of these always hold within an honest equilibrium, but they can break down elsewhere. Here we highlight cases where these intuitive results do not hold, and we then summarize with a complete description of the comparative statics.

More Competent Experts. Adding more competent experts always leads to better decisions (on average), but there are two scenarios where adding more competent experts (increasing p_g) can lead to more admission of uncertainty.

First, in the region where the bad types always guess, the good types are the only experts who send $m_\varnothing$, so increasing p_g leads to more admission of uncertainty.

More subtly, when the bad types sometimes send $m_\varnothing$ and $p_e < 1/2$ (the bottom left triangle), adding more competent experts leads the bad types to admit uncertainty more often. And since this is the part of the parameter space where good types usually admit uncertainty as well, adding more competent types leads to more admission of uncertainty overall:

Proposition 5. *In the equilibrium identified by proposition 4, (i) the expected value of the decision is strictly increasing in* p_g *and (ii) the unconditional probability that the expert admits uncertainty is strictly decreasing in* p_g *if* $p_e \leq p_g/\left(1+p_g\right)$ *(honest equilibrium) or if* $p_e \in \left[1/2, 1/\left(1+p_g\right)\right]$*, and it is strictly increasing in* p_g *otherwise.*

More Easy Questions. When the probability of an easy problem increases, this always decreases the admission of uncertainty because good types are more likely to be informed and bad types are more apt to guess. The potentially counterintuitive result here is that more easy problems can sometimes lead to worse decisions. This is possible because when problems are more likely to be easy, the bad types are more tempted to guess. If the bad types never actually guess (bottom right corner) or always guess (top right corner), this does not matter. However, when making problems easier actually makes the bad types guess more, the messages m_0 and m_1 become less informative. As shown in the following proposition, this can sometimes lead to worse decisions:

Proposition 6. *In the equilibrium identified by proposition 4, (i) the unconditional probability that an expert admits uncertainty is strictly decreasing in* p_e *and (ii) for any* p_g, *there exists a* $\widetilde{p}_e \in \left[p_g/\left(1+p_g\right), 1/\left(1+p_g\right)\right]$ *such that* v^* *is strictly decreasing in* p_e *for* $p_e \in \left[p_g/\left(1+p_g\right), \widetilde{p}_e\right]$.

DISCUSSION

This paper studies the strategic communication of uncertainty by experts with reputation concerns. Our analysis is built on two theoretical innovations. First, in our setup, the decision maker is uncertain not only about the state of the world, but also about whether the state is even knowable for qualified experts. This feature is related to prior work on the classification of uncertainty, specifically, “aleatory" and “epistemic" uncertainty, language dating to Hacking (1975). Aleatory uncertainty characterizes what we cannot know—“difficult" questions (e.g., the roll of dice)—where epistemic uncertainty is what we do not know but could if we were more informed.[16] We show that these properties of uncertainty have real implications both for understanding why communication about uncertainty is hard and for learning how to overcome that challenge. The way we formalize the distinction between easy and hard problems highlights the idea that part of being a domain expert is not merely knowing the answers to questions but knowing the limits of what questions are answerable.

A second innovation concerns the notion of “credible beliefs,” which is closely tied to structural consistency of beliefs (Kreps and Wilson 1982). Honest communication in our model is disciplined by experts' reputational concerns—off-path, they are punished by the low opinion of the decision maker. But what can she credibly threaten to believe? Our use of Markov sequential equilibrium rules out non-credible beliefs that stipulate updating on payoff-irrelevant information.

A pragmatic way to frame our inquiry is that we ask what would the decision maker want to learn, ex post to induce the experts to communicate their information honestly ex ante? We found that the intuitive answer—checking experts' reports against the true state of the world—is insufficient. Even when decision makers catch an expert red-handed in a lie, the severity of their beliefs is curtailed by the fact that good experts facing unanswerable questions are in the same conundrum as bad experts. Therefore, we show that state validation alone never induces honesty. In order to elicit honest reports from experts, it is necessary that the decision maker also learns whether the problem is difficult. Indeed, in environments where the expert has even very small policy concerns, difficulty validation alone may be sufficient.

Is such difficulty validation common in the real world? As discussed throughout, we believe sometimes information about difficulty naturally comes ex post, and it can sometimes be accomplished by methods like peer review. We conclude with a practical suggestion: that when consulting multiple experts, decision makers may want to give heterogenous incentives and ask different questions. Rewarding some for “getting things right” gives good incentives to avoid letting personal biases contaminate advice and potentially for collecting information in order to be informed in the first place. However, as emphasized here, these kinds of reward schemes may exacerbate the problem of getting experts to admit uncertainty. On the other hand, paying a flat fee to an expert who will likely not be consulted again for their services may have drawbacks, but it will make the expert much more comfortable admitting uncertainty. Further, some experts can simply be asked “do you think the evidence about this question is solid” rather than emphasizing what the expert thinks the truth is. Finding other ways to achieve difficulty validation could be a path to improving communication in politics and organizations more generally.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0003055420000209.

REFERENCES

Ashworth, Scott. 2012. “Electoral Accountability: Recent Theoretical and Empirical Work.” *Annual Review of Political Science* 15: 183–201.

Austen-Smith, David. 1990. Information Transmission in Debate. *American Journal of Political Science* 34 (1): 124–52.

Banks, Jeffrey S. 1990. “A Model of Electoral Competition with Incomplete Information.” *Journal of Economic Theory* 50 (2): 309–25.

Bergemann, Dirk, and Ulrich Hege. 2005. “The Financing of Innovation: Learning and Stopping.” *RAND Journal of Economics* 36 (4): 719–52.

Bergemann, Dirk, and Johannes Hörner. 2010. “Should Auctions Be Transparent?” *Cowles Foundation Discussion* Paper No. 1764.

Bhaskar, V., George J. Mailath, and Stephen Morris. 2013. “A Foundation for Markov Equilibria in Sequential Games with Finite Social Memory.” *Review of Economic Studies* 80 (3): 925–48.

Callander, Steven. 2011. “Searching for Good Policies.” *American Political Science Review* 105 (4): 643–62.

Calvert, Randall L. 1985. The Value of Biased Information: A Rational Choice Model of Political Advice.” *The Journal of Politics* 47 (2): 530–55.

Canes-Wrone, Brandice, Michael C. Herron, and Kenneth W. Shotts. 2001. Leadership and Pandering: A theory of Executive Policymaking.” *American Journal of Political Science* 45 (3): 532–50.

Cho, In-Koo, and David M. Kreps. 1987. “Signaling Games and Stable Equilibria.” *The Quarterly Journal of Economics* 102 (2): 179–221.

Crawford, Vincent P., and Joel Sobel. 1982. “Strategic Information Transmission.” *Econometrica: Journal of the Econometric Society* 50 (6): 1431–51.

Deb, Rahul, Mallesh Pai, and Maher Said. 2018. “Evaluating Strategic Forecasters.” *American Economic Review* 8 (10): 3057–103.

[16] The language of aleatory and epistemic uncertainty has taken particular hold in structural engineering, where it is important in the classification of risk; see Kiureghian and Ditlevsen (2009).

Dessein, Wouter. 2002. "Authority and Communication in Organizations." *The Review of Economic Studies* 69 (4): 811–38.

Dye, Ronald A. 1985. "Disclosure of Nonproprietary Information." *Journal of Accounting Research* 23 (1): 123–45.

Ericson, Richard, and Ariel Pakes. 1995. "Markov-perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies* 62 (1): 53–82.

Fearon, James D. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance." In *Democracy, Accountability, and Representation*, eds. A. Przeworski, S. C. Stokes, and B. Manin. Cambridge: Cambridge University Press, 55–97.

Fehrler, Sebastian, and Moritz Janas. 2019. Delegation to a Group. Unpublished Manuscript.

Fox, Justin, and Kenneth W. Shotts. 2009. Delegates or Trustees? A Theory of Political Accountability." *The Journal of Politics* 71 (4): 1225–37.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge, MA: MIT Press.

Gailmard, Sean, and John W. Patty. 2012. "Formal Models of Bureaucracy." *Annual Review of Political Science* 15: 353–77.

Gailmard, Sean, and John W. Patty. 2013. "Stovepiping." *Journal of Theoretical Politics* 25 (3): 388–411.

Gilligan, Thomas W., and Keith Krehbiel. 1987. "Collective Decisionmaking and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures." *Journal of Law, Economics, & Organization* 3 (2): 287–335.

Hacking, Ian. 1975. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability Induction and Statistical Inference*. Cambridge, UK: Cambridge University Press.

Harsanyi, John C. 1973. "Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-strategy." *International Journal of Game Theory* 2 (1): 1–23.

Harsanyi, John C., and Reinhard Selten. 1988. *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Holmström, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *The Review of Economic Studies* 66 (1): 169–82.

Hübert, Ryan. 2019. "Getting Their Way: Bias and Deference to Trial Courts." *American Journal of Political Science* 63 (3): 706–18.

Jung, Woon-Oh, and Young K. Kwon. 1988. "Disclosure When the Market is Unsure of the Information Endowment of Managers." *Journal of Accounting Research* 26 (1): 146–53.

Kartik, Navin. 2009. "Strategic Communication with Lying Costs." *Review of Economic Studies* 76: 1359–95.

Kiureghian, Armen Der, and Ove Ditlevsen. 2009. "Aleatory or Epistemic? Does It Matter?" *Structural Safety* 31: 105–12.

Kreps, David M., and Garey Ramey. 1987. "Structural Consistency, Consistency, and Sequential Rationality." *Econometrica* 55 (6): 1331–48.

Kreps, David M., and Robert Wilson. 1982. "Sequential Equilibria." *Econometrica* 50 (4): 863–94.

Kydd, Andrew. 2003. "Which Side Are You On? Bias, Credibility, and Mediation." *American Journal of Political Science* 47 (4): 597–611.

Leaver, Clare. 2009. Bureaucratic Minimal Squawk Behavior: Theory and Evidence from Regulatory Agencies." *American Economic Review* 99 (3): 572–607.

Little, Andrew T. 2017. "Propaganda and Credulity." *Games and Economic Behavior* 102: 224–32.

Manski, Charles F. 2019. Communicating Uncertainty in Policy Analysis. *Proceedings of the National Academy of Sciences* 116 (16):7634–41.

Maskin, Eric, and Jean Tirole. 1988a. "A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs." *Econometrica* 56 (3): 549–69.

Maskin, Eric, and Jean Tirole. 1988b. "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles." *Econometrica* 56 (3): 5 71–599.

Maskin, Eric, and Jean Tirole. 2001. "Markov Perfect Equilibrium: 1. Observable Actions." *Journal of Economic Theory* 100: 191–219.

Nash, John F. 1950. "The Bargaining Problem." *Econometrica* 18 (2): 155–62.

Ottaviani, Marco, and Peter Norman Sørensen. 2006. "Reputational Cheap Talk. RAND." *Journal of Economics* 37 (1).

Patty, John W. 2009. "The Politics of Biased Information." *The Journal of Politics* 71 (2): 385–97.

Prendergast, Canice, and Lars Stole. 1996. "Impetuous Youngsters and Jaded Old-timers: Acquiring a Reputation for Learning." *Journal of Political Economy* 104 (6): 1105–34.

Rappaport, Daniel. 2015. "Humility in Experts and the Revelation of Uncertainty." Unpublished Manuscript.

Schnakenberg, Keith E. 2015. "Expert Advice to a Voting Body." *Journal of Economic Theory* 160:102–13.

Schnakenberg, Keith E. 2017. "Informational Lobbying and Legislative Voting." *American Journal of Political Science* 61 (1): 129–45.

Tadelis, Steven. 2013. *Game Theory: An Introduction*. Princeton, NJ: Princeton University Press.

APPENDIX A: MARKOV SEQUENTIAL EQUILIBRIUM

General Definition of MSE

Take a general sequential game of incomplete information, and let each node (history) be associated with information set I and an action set A_I. Beliefs μ map information sets into a probability distribution over their constituent nodes. A strategy profile for the entire game σ maps each information set into a probability distribution over A_I. Write the probability (or density) of action a at information set I as $\sigma_I(a)$. Let the function $u_I(a, \sigma)$ denote the von Neumann–Morgenstern expected utility from taking action $a \in A_I$ at an information set I when all subsequent play, by all splayers, is according to σ. In our setting, the payoff-relevant state depends on the information set of the DM, $\mathcal{I}_{DM2}$, so to define it, we look to affine payoff equivalence following Harsanyi and Selten (1988) and Fudenberg and Tirole (1991).

Definition 2. *A strategy σ is a Markov strategy if whenever, for any pair of information sets I and I′ with associated action sets A_I and $A_{I'}$, and for some constants $\alpha > 0$ and β, there exists a bijection $f : A_I \rightarrow A_{I'}$ such that $u_I(a,\sigma) = \alpha u_{I'}[f(a),\sigma] + \beta$, $\forall a \in A_I$, then $\sigma_I(a) = \sigma_{I'}[f(a)]$.*

The extension of equilibrium in Markov strategies to a setting with incomplete information requires some additional language. Our notation and terminology parallels the treatment of sequential equilibrium in Tadelis (2013). As consistency is to sequential equilibrium, so Markov consistency is to Markov sequential equilibrium.

Definition 3. *A profile of strategies σ and a system of beliefs μ is Markov consistent if there exists a sequence of non-degenerate, Markov mixed strategies $\{\sigma^k\}_{k=1}^{\infty}$ and a sequence of beliefs $\{\mu^k\}_{k=1}^{\infty}$ that are derived from Bayes' Rule, such that $\lim_{k\to\infty}(\sigma^k, \mu^k) \rightarrow (\sigma, \mu)$.*

With this in hand, a notion of Markov sequential equilibrium follows directly.

Definition 4. *A profile of strategies* σ, *together with a set of beliefs* μ, *is a Markov sequential equilibrium if* (σ^*, μ^*) *is a Markov consistent perfect Bayesian equilibrium.*

Behavioral Motivation. Markov strategies have axiomatic foundations (Harsanyi and Selten 1988), and can be motivated by purification arguments as well as finite memory in forecasting (Bhaskar et al. 2013; Maskin and Tirole 2001). In the complete information settings to which it is commonly applied, the Markovian restriction prevents the players from conditioning their behavior on payoff-irrelevant aspects of the (common knowledge) history.[17] The natural extension of this idea to asymmetric information games is to prevent players from conditioning their strategies on payoff-irrelevant private information.

Our restriction on beliefs is also related to the notion of structural consistency proposed by Kreps and Wilson (1982).[18] In that spirit, Markov consistency formalizes a notion of "credible" beliefs, analogous to the notion of credible threats in subgame perfect equilibrium. Instead of using arbitrarily punitive off-path beliefs to discipline on-path behavior, we require that off-path beliefs are credible in the sense that, ex post, on arriving at such an off-path node, the relevant agent could construct a convergent sequence of Markov strategies to rationalize them.

Robustness of MSE. Though the restriction to Markov strategies itself enforces a notion of robustness, there is a trivial sense in which the restriction of Markov equilibrium—whether in a complete information setting or an incomplete information setting—is non-robust. In particular, because it imposes symmetric strategies only when incentives are exactly symmetric, small perturbations of a model may permit much larger sets of equilibria. In the standard applications of the Markov restriction, this could be driven by future payoffs being slightly different depending on the history of play. In our setting, good and bad uninformed experts could have marginally different expected payoffs. Either way, we maintain that this is a red herring. The point of the refinement, like the symmetry condition of Nash (1950), is to hold the theorist to a simple standard: that we be precise about exactly what *kind* of asymmetry in the model construction explains asymmetries in the predicted behavior. From this perspective, another interpretation of our work is that we are reflecting on exactly what informational structures introduce the asymmetry we need to obtain honesty in equilibrium. [19]

Note on a Dynamic Interpretation. We have formulated our game as a one-shot sequential game with reputational concerns, so the payoff equivalence holds without the need for affine transformations. In the repeated game implied by the reputational concerns, this will not generally be the case. Depending on the formulation of payoffs, good experts will most likely have a higher continuation value than bad ones in all but a babbling equilibrium. This is where the potential for affine transformations in our definition of Markov strategies is useful—if the prior beliefs of the DM at the beginning of a period are sufficient for the history of the game, then setting β equal to the difference in continuation values means that the Markov restriction still binds.

MSE and SE

Here we offer a brief discussion of the prior use of the Markov sequential equilibrium (MSE) solution concept as well as an illustration of its implications as a refinement on off-path beliefs.

MSE is the natural extension of Markov perfect equilibrium to incomplete information games. However, its usage is infrequent and sometimes informal. To our knowledge, there is no general treatment or general guidance to the construction of the maximally coarse (Markov) partition of the action space, unlike the case of MPE (Maskin and Tirole 2001). Bergemann and Hege (2005) and Bergemann and Hörner (2010) employ the solution concept, defining it as a perfect Bayesian equilibrium in Markovian strategies. In other words, they impose the Markov restriction only on the sequential rationality condition. This is different and weaker than our construction. Our definition of MSE imposes the Markov assumption on both sequential rationality as well as consistency. While they do not use the Markov restriction to refine off-path beliefs, this is of no consequence for their applications.

To see the relevance of MSE to off-path beliefs, consider the game illustrated in Figure A.1, which is constructed to mirror an example from Kreps and Wilson (1982).[20] First, Nature chooses Player 1's type, *a* or *b*. Next, Player 1 chooses *l* or *r*. Finally, Player 2 chooses *u* or *d*. Player 2 is never informed of Player 1's type. Whether Player 1 knows their own type is the key difference between the two games.

[17] Applications of Markov equilibrium have been similarly focused on the infinitely repeated, complete information setting. See, e.g., Maskin and Tirole (1988a, 1988b) and Ericson and Pakes (1995).

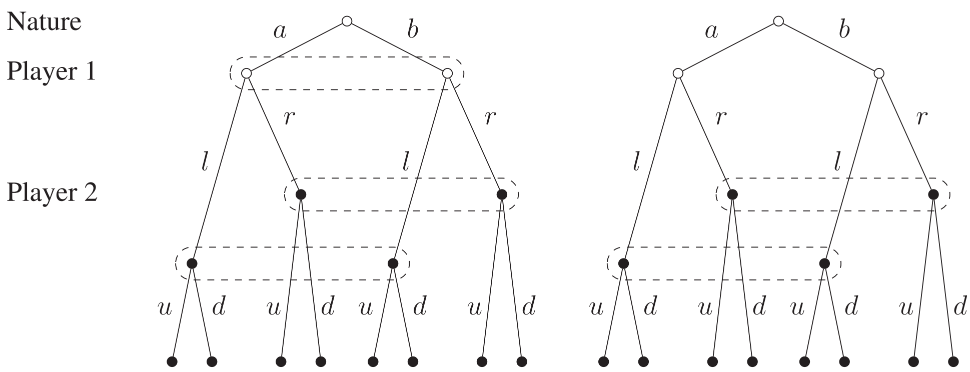
[18] Kreps and Ramey (1987) demonstrated that consistency may not imply structural consistency, as conjectured by Kreps and Wilson (1982). We observe that as the Markov property is preserved by limits, Markov consistency does not introduce any further interpretive difficulty.

[19] We thank an anonymous referee for pointing out that one could also develop a notion of ε – Markov equilibrium to make this point. This is beyond the theoretical ambition of the current work, but an interesting direction for future work.

[20] See, in particular, their Figure 5 (page 873).

https://doi.org/10.1017/S0003055420000209 Published online by Cambridge University Press

FIGURE A.1. Consistency, Markov Consistency, and Off-Path Beliefs



Note: This figure depicts two games, which differ in whether Player 1 knows their own type. Their type, *a* or *b*, is chosen by Nature with *Pr*(a)= p and *Pr*(b) = 1−p. Player 1 chooses *l* or *r*, and Player 2 sees this and reacts with *u* or *d*. Payoffs are omitted, but they can be written $u_i(\cdots)$.

In the first game, the player does not know their type. Posit an equilibrium in which Player 1 always chooses *l*. What must Player 2 believe at a node following *r*? If the theorist is studying perfect Bayesian equilibrium (PBE), they may specify any beliefs they wish. Alternatively, if they are studying sequential equilibrium (SE), Player 2 must believe that Player 1 is of type *a* with probability *p*.

In the second game depicted, SE imposes no restriction on Player 2's off-path beliefs. However, MSE may. If $u_1(a,l,\cdot)=u_1(b,l,\cdot)$ and $u_1(a,r,\cdot)=u_1(b,r,\cdot)$ (or, more generally, the expected utilities are equal up to an affine transformation), then we say that Player 1's type is *payoff irrelevant.* The restriction to Markov strategies implies that Player 1's strategy does not depend upon their type. Markov consistency implies that, further, Player 2 cannot update about payoff irrelevant information. Therefore Player 2 must believe that Player 1 is of type *a* with probability *p*.

Non-Markovian PBE

Here we briefly discuss PBEs that fail the Markov consistency requirement of MSE, and we argue why we believe these equilibria are less sensible.

In particular, we demonstrate that the most informative equilibrium under no policy concerns can involve more transmission of uncertainty and also information about the state. However, these equilibria are not robust to minor perturbations, such as introducing a vanishingly small random cost of lying.

Example 1: Admission of Uncertainty with No Validation. Even without the Markov restriction, it is immediate that there can be no fully honest equilibrium with no validation. In such an equilibrium, the competence assessment for sending either m_0 or m_1 is 1, and the competence assessment for sending $m_\varnothing$ is strictly less than one. So the uninformed types have a strict incentive to deviate to m_0 or m_1. However, unlike the case with the Markov restriction that leads to babbling, there is an always guessing equilibrium: If informed types always send $m_x=s_x$ and all uninformed types send m_1 with probability p_1 and m_0 otherwise, the competence assessment upon observing either message is p_g. So no type has an incentive to deviate.

Further, it is possible to get admission of uncertainty if the good and bad uninformed types play different strategies. In the extreme, suppose the good types always send their honest message, including the uninformed sending $m_\varnothing$. If the bad types were to always send m_0 or m_1, then the competence assessment upon sending $m_\varnothing$ would be 1. In this case, saying "I don't know" would lead to the highest possible competence evaluation, giving an incentive for all to admit uncertainty even if they know the state.

It is straightforward to check that if the bad types mix over messages $(m_0, m_1, m_\varnothing)$ with probabilities $[(1-p_1), p_e p_1, 1-p_e]$, then the competence assessment upon observing all messages is p_g, and so no expert has an incentive to deviate.

A common element of these equilibria is that the competence assessment for any on-path message is equal to the prior. In fact, a messaging strategy can be part of a PBE if and only if this property holds: the competence assessments must be the same to prevent deviation, and if they are the same, then by the law of iterated expectations they must equal the prior. So, there is a range of informative equilibria, but they depend on types at payoff-equivalent information sets taking different actions, a violation of Markov strategies that reflects their sensitivity to small perturbations of the payoffs.

Example 2: Honesty with State Validation or Difficulty Validation. Now return to the state validation case and the conditions for an honest equilibrium. Without the Markov restriction on beliefs, it is possible to set the off-path belief upon observing an incorrect guess to 0. With this off-path belief, the incentive compatibility constraint to prevent sending m_1 becomes $\pi_g^{\varnothing} \geq p_1$. Since $\pi_g^{\varnothing}$ is a function of p_g and p_e (but not p_1), this inequality holds for a range of the parameter space. However, this requires beliefs that are not Markov consistent—the DM who reaches that off-path node cannot construct a Markov strategy to rationalize their beliefs. So we argue that the threat of these beliefs is not credible. Similarly, without the Markov restriction it is possible to get honesty with just difficulty validation. The binding constraint is that if any off-path message leads to a zero competence evaluation, the bad type gets a higher payoff from sending $m_{\varnothing}$—as in the case with $\gamma \to 0$, $(1-p_e)p_g$—than from sending m_1 (now p_e). So, honesty is possible if $(1-p_e)p_g > p_e$—the same condition as when $\gamma \to 0$.

The Fragility of These Examples. A standard defense of Markov strategies in repeated games is that they represent the simplest possible rational strategies (Maskin and Tirole 2001). A similar principle applies here: rather than allowing for types with the same (effective) information to use different mixed strategies sustained by indifference, MSE focuses on the simpler case where those with the same incentives play the same strategy.

Further, as shown by Bhaskar, Mailath, and Morris (2013) for the case of finite social memory, taking limits of vanishing, independent perturbations to the payoffs—in the spirit of Harsanyi and Selten (1988) "purification"—results in Markov strategies as well. Intuitively, suppose the expert receives a small perturbation to his payoff for sending each message that is independent of type and drawn from a continuous distribution, so he has a strict preference for sending one message over the others with probability one. Payoff-indifferent types must use the same mapping between the perturbations and messages, analogous to Markovian strategies. Further, if these perturbations put all messages on path, then all beliefs are generated by Markovian strategies.[21]

Summary. It would be possible to construct additional informative equilibria if we allowed different types to play different actions, even when they are payoff equivalent. We view this as a modeling contrivance, and this is precisely what the Markov consistency restriction, above and beyond standard consistency, restricts. This point was previously made by Harsanyi and Selten (1988), who contend that the property of "invariance with respect to isomorphisms," on which our definition of Markov strategies is based, is "an indispensable requirement for any rational theory of equilibrium point selection that is based on strategic considerations exclusively." Or, in the appeal of Maskin and Tirole (2001) to payoff perturbations, "minor causes should have minor effects."

APPENDIX B: PROOFS OF RESULTS IN THE MAIN TEXT

More General Definitions. Some of our results in this section rely on more general definitions of messaging strategies. Starting with "honesty":

Definition 5. *Let $\pi_s(m)$ be the DM posterior belief that the expert observed signal s upon sending message m. An equilibrium is* ***honest*** *if $\pi_s(m) \in \{0,1\}$ $\forall\, s \in S$ and all on-path m.*

As in all cheap-talk games, the messages sent only convey meaning by which types send them in equilibrium. We define admitting uncertainty as sending a message which is never sent by either informed type:

Definition 6. *Let M_0 be the set of messages sent by the s_0 types and M_1 be the set of message sent by the s_1 types. Then an expert* ***admits uncertainty*** *if he sends a message $m \notin M_0 \cup M_1$.*

Finally, an important class of equilibria will be one in which the informed types send distinct message from each other, but the uninformed types sometimes if not always mimic these messages:

Definition 7. *A* ***guessing*** *equilibrium is one where $M_0 \cap M_1 = \varnothing$, and $Pr(m \in M_0 \cup M_1|\theta,s_{\varnothing}) > 0$ for at least one $\theta \in \{g,b\}$. In an* ***always guessing*** *equilibrium, $Pr(m \in M_0 \cup M_1|\theta,s_{\varnothing}) = 1$ for both $\theta \in \{g,b\}$.*

That is, an always guessing equilibrium is one where the informed types report their signal honestly, but the uninformed types never admit uncertainty.

Proof of Proposition 1: For convenience, we extend the definition of v so $v(a,\pi_1)$ represents the expected quality of policy a under the belief that the state is 1 with probability π_1.

The DM's expected payoff from the game can be written as the sum over the (expected) payoff as a function of the expert signal:

$$\sum_{s\,\in\,\{s_0,s_1,s_{\varnothing}\}} Pr(s) \sum_{m} Pr(m|s) v[a^*(m), Pr(\omega|s)]. \tag{7}$$

In the honest equilibrium, when the expert observes s_0 or s_1, the DM takes an action equal to the state with probability 1, giving payoff 1. When the expert observes $s_{\varnothing}$, the equilibrium action is p_1 giving payoff $v(p_1,p_1) = 1 - p_1(1-p_1)$. So, the average payoff is

[21] A related refinement more specific to our setting is to allow for a small random "lying cost" for sending a message not corresponding to the signal, which is independent of the type (Kartik 2009).

https://doi.org/10.1017/S0003055420000209 Published online by Cambridge University Press

$$p_g p_e 1 + \left(1 - p_g p_e\right) p_1 (1 - p_1) = \bar{v}.$$

This payoff as expressed in Equation 7 is additively separable in the signals, and v is strictly concave in a for each s. So, for each $s \in \{s_0, s_1, s_\varnothing\}$, this component of the sum is maximized if and only if $a^*(m)$ is equal to the action taken upon observing the honest message is with probability 1. That is, it must be the case that

$$a^*(m) = \begin{cases} 1 & m : Pr(m|s_1) > 0 \\ p_1 & m : Pr(m|s_\varnothing) > 0 \\ 0 & m : Pr(m|s_0) > 0 \end{cases}. \tag{8}$$

If the equilibrium is not honest, then there must exist a message m' such that $Pr(s|m') < 1$ for all s. At least one of the informed types must send m' with positive probability; if not, $Pr(s_\varnothing|m') = 1$. Suppose the type observing s_0 sends m' with positive probability (An identical argument works if it is s_1). To prevent $Pr(s_0|m') = 1$ another type must send this message as well, and so in response the DM chooses an action strictly greater than 0, contradicting the condition in Equation 8 and hence the expected quality of the decision in any equilibrium that is not honest is strictly less than $\bar{v}$. □

Proof of Proposition 2: For any messaging strategy, the DM must form a belief about the expert competence for any message (on- or off-path); since it only depends on this message (and not the validation as in other cases) write these $\pi_g(m)$. So, for any type θ, the expected utility for sending message m is just $\pi_g(m)$. All types are payoff-equivalent in any equilibrium, so in any MSE they must use the same strategy. Since all messages are sent by both informed and uninformed types, there is no admission of uncertainty. □

Proof of Proposition 3: A more complete description of the MSE with no policy concerns and state validation is

Proposition 7. *With state validation and no policy concerns (i) in any MSE, there is no admission of uncertainty and (ii) any non-babbling MSE is equivalent, subject to relabeling, to an MSE where both uninformed types send m_1 with probability*

$$\sigma^*_\varnothing(m_1) = \begin{cases} \dfrac{p_1\left(1 + p_g p_e\right) - p_g p_e}{1 - p_g p_e} & \text{if } p_1 < 1/\left(1 + p_g p_e\right) \\ 1 & \text{otherwise,} \end{cases}$$

and m_0 with probability $\sigma^*_\varnothing(m_0) = 1 - \sigma^*_\varnothing(m_1)$.

Proof. Part i is immediate in a babbling equilibrium: there is no admission of uncertainty since there are no messages only sent by the uninformed types. Propositions S.7 and S.8 in section 2 of the Supplemental Information show that with state validation and no policy concerns, all non-babbling MSE are equivalent, subject to a relabeling of the messages, to one where the the s_0 types send m_0, the s_1 types send m_1, and there is only one other potential message $m_\varnothing$. What remains to be shown is that in the MSE of this form, the uninformed types never send $m_\varnothing$ and send m_0 and m_1 with the probabilities in the statement of the proposition.

Recall the Markov strategy restriction implies the good and bad uninformed types use the same strategy. As shown in the main text, in a conjectured honest equilibrium, the payoff for an uninformed type to send $m_\varnothing$ is $\pi_g^\varnothing$. To formally show a deviation to m_1 is profitable, recall the payoff to sending this message when $\omega = 1$ is 1. When $\omega = 0$, the competence assessment is off path. Markov consistency requires that this belief be formed as the limit to a sequence of well-defined beliefs that are consistent with a corresponding sequence of Markov strategies. Take any sequence of Markov strategies σ^k and resulting beliefs

$$\pi_g^k(m_1, 0) = \frac{Pr(m_1, 0, \theta = g)}{Pr(m_1, 0)} = \frac{(1 - p_1) p_g p_e \sigma_0^k(m_1) + (1 - p_1) p_g (1 - p_e) \sigma_\varnothing^k(m_1)}{(1 - p_1) p_g p_e \sigma_0^k(m_1) + (1 - p_1)\left(1 - p_g p_e\right) \sigma_\varnothing^k(m_1)}.$$

This belief is increasing in $\sigma_0^k(m_1)$ and decreasing in $\sigma_\varnothing^k(m_1)$, and it can range from $\pi_g^\varnothing$—when $\sigma_0^k(m_1) = 0$ and $\sigma_\varnothing^k(m_1) > 0$—to 1—when $\sigma_0^k(m_1) > 0$ and $\sigma_\varnothing^k(m_1) = 0$. So, $\pi_g(m_1, 0)$ must be the limit of a sequence where each element lies in $\left[\pi_g^\varnothing, 1\right]$, and the off-path belief must lie on this interval as well. Hence, the payoff to deviating to m_1 is at least $p_1 + (1 - p_1)\pi_g^\varnothing > \pi_g^\varnothing$, completing the proof that there is no honest equilibrium.

Now suppose the uninformed types send $m_\varnothing$ with a probability strictly between 0 and 1. The competence assessment for sending $m_\varnothing$ is $\pi_g^\varnothing$. Writing the probability the uninformed types send m_1 with $\sigma_\varnothing(m_1)$, the competence assessment for sending m_1 and observing validation that $w = 1$ is

https://doi.org/10.1017/S0003055420000209 Published online by Cambridge University Press

$$\begin{aligned} Pr(\theta=g|m_1;\sigma_\varnothing(m_1)) &= \frac{p_g p_e p_1 + p_g(1-p_e)\sigma_\varnothing(m_1)}{p_g p_e p_1 + \left[p_g(1-p_e)+\left(1-p_g\right)\right]\sigma_\varnothing(m_1)} \\ &\ge \frac{p_g p_e p_1 + p_g(1-p_e)}{p_g p_e p_1 + \left[p_g(1-p_e)+\left(1-p_g\right)\right]} \\ &> \frac{p_g(1-p_e)}{\left[p_g(1-p_e)+\left(1-p_g\right)\right]} = \pi_g^\varnothing . \end{aligned}$$

Since the competence assessment for sending m_1 is strictly higher than for sending $m_\varnothing$, there can be no MSE where the uninformed types admit uncertainty, completing part i.

For part ii, first consider the condition for an equilibrium where both m_0 and m_1 are sent by the uninformed types. The uninformed types must be indifferent between guessing m_0 and m_1. This requires

$$p_1\pi_g(m_1,\omega=1)+(1-p_1)\pi_g^\varnothing=(1-p_1)\pi_g(m_0,\omega=0)+p_1\pi_g^\varnothing, \tag{9}$$

where the posterior beliefs about competence when “guessing wrong” are $\pi_g^\varnothing$ and when “guessing right” are given by Bayes’ rule:

$$\begin{aligned} \pi_g(m_1,\omega=1) &= \frac{Pr(\theta=g,\omega=1,m_1)}{Pr(m_1,\omega=1)} = \frac{p_1 p_g[p_e+(1-p_e)\sigma_\varnothing(m_1)]}{p_1\left[p_g p_e+\left(1-p_g p_e\right)\sigma_\varnothing(m_1)\right]} \\ \pi_g(m_0,\omega=0) &= \frac{Pr(\theta=g,\omega=0,m_0)}{Pr(m_0,\omega=0)} = \frac{(1-p_1)p_g[p_e+(1-p_e)\sigma_\varnothing(m_0)]}{(1-p_1)\left[p_g p_e+\left(1-p_g p_e\right)\sigma_\varnothing(m_0)\right]}. \end{aligned}$$

Plugging these into Equation 9 and solving for the strategies with the additional constraint that $\sigma_\varnothing(m_0)+\sigma_\varnothing(m_1)=1$ gives

$$\begin{aligned} \sigma_\varnothing(m_0) &= \frac{1-p_1\left(1+p_g p_e\right)}{1-p_g p_e} \\ \sigma_\varnothing(m_1) &= \frac{p_1\left(1+p_g p_e\right)-p_g p_e}{1-p_g p_e}. \end{aligned}$$

For this to be a valid mixed strategy, it must be the case that both of these expressions are between zero and one, which is true if and only if $p_1 < 1/\left(1+p_g p_e\right) \in (1/2,1)$. So, if this inequality holds and the off-path beliefs upon observing $m_\varnothing$ are sufficiently low, there is an MSE where both messages are sent by the uninformed types. And the competence assessment for any off-path message/validation can be set to $\pi_g^\varnothing$ (i.e., the lowest belief possible with Markov consistency), which is less than the expected competence payoff for sending either m_0 or m_1.

Now consider an equilibrium where uninformed types always send m_1. The on-path message/validation combinations are then $(m_1,\omega=0)$, $(m_1,\omega=1)$, and $(m_0,\omega=0)$, with the following beliefs about the expert competence:

$$\begin{aligned} \pi_g(m_1,\omega=0) &= \frac{p_g(1-p_e)}{p_g(1-p_e)+1-p_g}; \\ \pi_g(m_1,\omega=1) &= \frac{p_g p_e+p_g(1-p_e)}{p_g p_e+p_g(1-p_e)+\left(1-p_g\right)} = p_g, \text{ and} \\ \pi_g(m_0,\omega=0) &= 1. \end{aligned}$$

Preventing the uninformed types from sending m_0 requires

$$p_1 p_g+(1-p_1)\frac{p_g(1-p_e)}{p_g(1-p_e)+1-p_g} \ge p_1\pi_g(m_0,\omega=1)+(1-p_1).$$

This inequality is easiest to maintain when $\pi_g(m_0,\omega=1)$ is small, and by the argument in the main text, in an MSE it must be at least $\pi_g^\varnothing$. Setting $\pi_g(m_0,\omega=1)=\pi_g^\varnothing$ and simplifying gives $p_1 \ge 1/\left(1+p_g p_e\right)$, which is the reverse of the

https://doi.org/10.1017/S0003055420000209 Published online by Cambridge University Press

inequality required for an MSE where both m_0 and m_1 are sent. Again, setting the competence assessment for an off-path message to $\pi_g^{\varnothing}$ prevents this deviation.

So, if $p_1 \leq 1/\left(1+p_g p_e\right)$ there is an MSE where both messages are sent, and if not there is an MSE where only m_1 is sent.

Finally, it is easy to verify there is never an MSE where only m_0 is sent, as the uninformed types have an incentive to switch to m_1. □

Proof of Proposition 4: See the Proof of Proposition S.5 in section 1 of the Supplemental Information.

Proof of Proposition 5: For part i, the result is immediate in the range of p_e where p_g does not change the bad type strategy. For the range where the bad type strategy is a function of p_g, plugging in the strategies identified into Equation 6 and simplifying gives the expected quality of the decision as

$$1 - p_1(1-p_1) + \frac{\left(p_e p_g\right)^2 p_1(1-p_1)}{p_e - p_g(1-2p_e)}. \tag{10}$$

The derivative of Equation 10 with respect to p_g is

$$\frac{p_1(1-p_1)p_e^2 p_g\left[2p_e\left(1+p_g\right) - p_g\right]}{\left[p_e - p_g(1-2p_e)\right]^2},$$

which is strictly positive if $p_e > \frac{p_g}{2\left(1+p_g\right)}$. Since the range of p_e where the bad type plays a mixed strategy is $p_e \in \left[p_g/\left(1+p_g\right), 1/\left(1+p_g\right)\right]$, this always holds.

For part ii, the unconditional probability of admitting uncertainty in this equilibrium is

$$p_g(1-p_e) + \left(1-p_g\right)\sigma_b(m_{\varnothing}). \tag{11}$$

Within the range for an honest equilibrium—where $\sigma_b(m_{\varnothing}) = 1$—the derivative of Equation 11 with respect to p_g is $-p_e < 0$. In the range where the bad type always guesses—$\sigma_b(m_{\varnothing}) = 0$—the derivative is $(1-p_e) > 0$. Plugging in the equilibrium strategy when interior and differentiating with respect to p_g gives $1-2p_e$, which is positive if and only if $p_e < 1/2$. □

Proof of Proposition 6: For part i, $\sigma_b(m_{\varnothing})$ is weakly decreasing in p_e, so Equation 11 is strictly decreasing in p_e. For part ii, in the range $p_e \in \left[p_g/\left(1+p_g\right), 1/\left(1+p_g\right)\right]$, the expected quality of the decision is Equation 10. Differentiating this with respect to p_e gives

$$\frac{p_1(1-p_1)p_e p_g^2\left(p_e - 2p_g + 2p_e p_g\right)}{\left(p_e - p_g + 2p_e p_g\right)^2},$$

which evaluated at $p_e = p_g/\left(1+p_g\right)$ simplifies to $-p_1(1-p_1)$. By continuity, this derivative must be negative on some nonempty interval $\left[p_g/\left(1+p_g\right), \tilde{p}_e\right]$. So, the value of the decision must be locally decreasing at $p_e = p_g/\left(1+p_g\right)$, and by continuity, for an open interval $p_e \in \left[p_g/\left(1+p_g\right), \tilde{p}_\delta\right]$. □