Discussion Paper Series – CRC TR 224

Discussion Paper No. 093
Project B 01

Fake Experts

Patrick Lahr*
Justus Winkelmann**

May 2019

*University of Bonn, patricklahr@uni-bonn.de
**University of Bonn, jwinkelmann@uni-bonn.de

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Fake Experts

Patrick Lahr and Justus Winkelmann*

May 14, 2019

Abstract

We consider a multi-sender cheap talk model, where the receiver faces uncertainty over whether senders have aligned or state-independent preferences. This uncertainty generates a trade-off between giving sufficient weight to the most informed aligned senders and minimizing the influence of the unaligned. We show that preference uncertainty diminishes the benefits from specialization, i.e., senders receiving signals with more dispersed accuracy. When preference uncertainty becomes large, it negates them entirely, causing qualified majority voting to become the optimal form of communication. Our results demonstrate how political polarization endangers the ability of society to reap the benefits of specialization in knowledge.

Keywords: Cheap Talk, Information Aggregation, Voting

*Lahr: University of Bonn, patricklahr@uni-bonn.de. Winkelmann: University of Bonn, jwinkelmann@uni-bonn.de. We thank Ian Ball, Felix Bierbrauer, Jan Knöpfle, Daniel Krähmer, Stephan Lauermann, Benny Moldovanu, Martin Pollrich, Larry Samuelson, Dezsö Szalay, and Aleh Tsyvinski. Both authors wish to express their gratitude for the hospitality of the Yale Economics department during their respective visits, amid which much of this research has been conducted. Funding by the Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224 is gratefully acknowledged.

1 Introduction

Consider a receiver, who must decide on an action, but who relies on multiple senders, to obtain relevant information.[1] Senders are imperfectly informed through private signals, which have more and less informative realizations. They have private preferences that are either aligned with the receiver's, we call those *advisory senders*, or are independent of the information relevant to the receiver, we call those *partisans*. What is the best way the receiver can communicate with senders, given that their preferences are unknown to her? How does the effect of this uncertainty depend on the information available to senders? Lastly, how does optimal communication relate to simple and frequently used tools to aggregate information, such as voting?

To illustrate these questions, let us consider the example of a judge trying to decide whether a defendant is guilty or innocent. After hearing multiple conflicting testimonies, she must decide the case, knowing that some witnesses, might give her false information, due to relations with the defendant or other suspects. If one person claims to have seen the crime, can the judge believe him, knowing that he might make such a claim to maximize his influence?

A similar situation arises when a government agency relies on expert advice for regulatory decisions, such as approving a new drug. In these cases, scientists hold information that is valuable in making the decision, yet they might also have private interests owing to financial ties to the pharmaceutical companies.[2] Due to different specializations and experiences with the particular drug, it is likely that the scientists in the advisory bodies are differently informed. In such a situation, should the regulator give more weight to those scientists who express high confidence in their positions? This allows her to account for the heterogeneity of information, but it also increases the possibility of scientists with conflicts of interest exerting greater influence by falsely claiming high confidence. Or should she simply decide by the numbers of individual votes for and against approval?

Lastly, how should individuals adjust their learning when they might be lied to by interested parties? This concern extends beyond commercial interest, and has recently been at the center of political debates, most notably that of "fake news." We will therefore try to understand the impact of a voter struggling to distinguish between interest-led rhetoric and reputable news. Will critical information still be heard, or will public discourse be

[1] Throughout the paper, we refer to the receiver with the female and the senders with the male pronoun.

[2] In an article in *Science*, Piller [2018] discusses the specific case of approval of a drug designed to prevent heart attacks and strokes. He finds that one scientist on the admission panel received more than $2 million for various purposes from the drug manufacturing pharmaceutical company. Further, Piller [2018] documents that the majority of 107 scientists who advised the Food and Drug Administration in the United States on the approval of 28 drugs from 2008 to 2014 received payments from makers of the drugs or competing firms.

reduced to deciding which side has gathered sufficient numbers?

In this paper, we examine such decision problems using a multi-sender cheap talk model in which senders receive conditionally independent information. In contrast to most models in the literature, ours does not assume that senders have a fixed and known bias. Instead, we assume that they have either aligned interests or state-independent preferences. This assumption works well to capture situations like the above, in which there is a clear common interest objective typically trumped by the private considerations of the senders (if there are any). As is standard in the literature, we focus throughout on the most informative equilibrium. We view the information transmitted in this equilibrium as a natural benchmark to voting games in which agents are restricted to two messages and the receiver has ex-ante committed herself to a threshold rule to translate messages into actions.

To understand the precise effect of partisanship in our model, we need to analyze senders' information structures. To illustrate this, let us briefly return to the example of the scientists advising a regulator on a new drug. Any information a scientist receives has both a direction, i.e., it favors either approval or rejection of the drug, and an intensity, i.e., how much it moves the scientist's belief away from the prior. We will call the mean intensity of a scientist's signal his *average informativeness*. Holding this average intensity constant, we will further differentiate between more specialized scientists that receive signals with very heterogeneous intensities and more generalized ones. In our example, a *specialized* scientist will learn whether the question is in his field of expertise and hence update his posterior more or less than his generalist colleague.

When interest is common, specialization has enormous benefits. One sender who happens to receive a perfect signal is more informative than five with fairly accurate signals. In the presence of partisanship, however, any differential treatment of messages based on the claimed intensity of the underlying signal also increases partisans' ability to pick messages that have more weight in the decision.

Equilibrium behavior in the presence of this uncertainty is characterized by senders with aligned interest stating their true beliefs, while partisans send messages independently of their information, mimicking senders that receive the most informative signals in their preferred direction. The receiver acts on the central trade-off to use as much information from aligned senders as possible while limiting the influence of partisans, and puts caps on the influence any one message can have on her decision. However, between these caps communication between aligned senders and the receiver remains perfect. This is a notable difference from the usual coarsening of messages in cheap talk games with known bias. As a consequence, partisanship has two distinct effects on information transmission, which relate to our two concepts of informational content.

The first effect is that the information held by partisans is lost, since they message independently of the signal they receive. This leads to a proportional loss in average informativeness. Such a loss is due to the mere existence of partisans and is equally effective if senders' preferences are known to the receiver; hence partisan messages can be ignored. It is also similar to viewing partisan types as incompetent and simply losing their signal.

The second effect is caused by the indistinguishability of advisory and partisan experts. Partisanship is thus created for the receiver and results in the uninformative messages of the partisans, being treated the same as the messages from those senders with the most valuable signals. This leads to a loss in effective specialization, since the best signals are now diluted by signals that are on average uninformative, making the message appear more homogeneous overall.

If partisanship becomes sufficiently large, all gains from specialization are wiped out and senders reduce their messages to a mere indication of the direction of their information, while the receiver bases her decision on whether the number of messages in favor of one alternative meets a fixed threshold. This binary communication between senders and receiver resembles a form of qualified majority voting. Consequently even in situations where it is natural to assume some senders have more to contribute than others arises as an optimal way of communication. In these environments, the average informativeness of each sender becomes the decisive predictor of the receiver's ability to match the state with her decision, while specialization becomes worthless. This can lead to a reversal where a group of expert advisors with more heterogeneously distributed posteriors is preferable to a receiver, when partisanship is low, but performs worse when partisanship is high. We are thus worried that increasing political polarization might substantially diminish the gains society can reap from increases in the specialization of knowledge.

Lastly, we discuss how we can generalize our analysis to the case where the receiver already has some information that distinguishes senders. She might be more or less sure about some senders' preferences and might also have information differentiating the type of information certain senders have access to. Again it becomes apparent that specialization is only valuable as long as it is accompanied by a high level of certainty over preference alignment.

The paper continues as follows. In the rest of this section, we review the literature. In section 2, we introduce our model. Section 3 analyzes the special case in which all senders have aligned interests. We use this natural benchmark to contrast our later findings. Section 4 introduces concepts to analyze both senders' information structure and the information that is transmitted to the receiver. We apply these concepts to our general model in section 5 and derive our main results. Section 6 concludes.

Related Literature

We place our paper between the literatures on cheap talk and information aggregation in voting. The former builds on the seminal work of Crawford and Sobel [1982] and analyzes strategic communication between a better-informed sender and a receiver whose action determines the payoff of both. In their original setup, the sender has private and perfect information on a one-dimensional state of the world and a bias known to the receiver.

We depart from this classical model in three central ways, with multiple senders, who are imperfectly informed, and whose preference are unknown to the sender.

Gilligan and Krehbiel [1989] have been the first to study a model with multiple senders. In their model two privately and perfectly informed senders with publicly-known biases communicate with a receiver. The focus of their analysis is the comparison of three communication protocols that comprise different forms of cheap talk. Similarly, Krishna and Morgan [2001] study a setting with two senders that sequentially send public messages to a receiver. The degree of information revelation depends on whether the senders have aligned or opposing biases.

Austen-Smith [1990] is the first to study a cheap talk problem in which senders are imperfectly informed about a binary state of the world. While he identifies circumstances under which a cheap talk phase alters the decision, Wolinsky [2002] solves for the most efficient communication structure.

In Alonso et al. [2008] and Hummel et al. [2013], uncertainty about senders' preferences arises endogenously, since each sender is interested in the decision matching his type, while the receiver wants to match the decision to the average of senders' types. Hence a sender's type contains both relevant information about the state, i.e., the average of types, as well as the bias, i.e., the distance of the individual sender's type from the average.

In contrast to this approach and in line with our own, Li and Madarász [2008] analyze a single-sender game with private bias that does not enter the receiver's payoff. They find that both players can benefit from the privacy of the sender's bias when it is not too large. This does not hold in our model, since we assume that private biases are such that senders' preferences become state-independent.

Lastly, the research closest to our own contribution is the independent working paper by Glazer et al. [2019] modeling internet recommendation systems. Their work, like ours, studies a model with multiple imperfectly informed senders; however, not all senders in their model are strategic. "Honest" senders always communicate their true signal, while only unaligned senders act strategically. Since we find in our model that the most informative equilibrium is one in which aligned senders strategically decide to be honest, the equilibria in both models are similar. However, given the different economic motivations of the two works, the analyses building on these equilibria are distinct.

The second body of literature we relate to is on information aggregation in voting. It goes back to Condorcet [1785] and his famous jury theorem, stating that large groups of independently informed senders select the correct alternative with near certainty. He assumes that senders vote sincerely, while Feddersen and Pesendorfer [1997] establish a similar result for strategic senders. They show that when the number of voters grows large, privately-held information leads to the same decision as public information.

Despite the effectiveness of voting for information aggregation in large populations, the same literature has brought up effects like the swing voter's curse, first discussed in Feddersen and Pesendorfer [1996], that illustrate a loss of information in small populations. This loss is mainly due to the nature of the voting game, with its limited amount of messages, usually two or three, and its fixed threshold. Any such voting rule can be interpreted in our model as a behavioral type of receiver to which strategic voters optimally react. We therefore believe, that our model with its strategic receiver, provides a natural benchmark for voting systems and helps to distinguish which losses of information are necessary consequences of conflicts of interest and which are due to the specific features of real-world voting systems.

Previously, McMurray [2017] and Azrieli [2018] have worked on the limits of elections with few available messages. McMurray [2017] studies a common interest election of ex-ante symmetric candidates by a fixed number of heterogeneously informed senders. In equilibrium, voters coordinate around specific candidates to transmit information. His model can be interpreted as a cheap talk game with a restricted number of messages. If the number of candidates becomes large, the model converges to our common interest setting. Azrieli [2018] analyzes the loss of anonymous voting rules if senders are publicly known to be differently well-informed. The common-value analysis is also closely related to ours. However, we assume that signals are private information and focus on their interplay with private interests.

2 The Model

There is a set of senders $\{1, ..., n\}$ and a receiver indexed by 0. Each sender i receives a signal about the unknown state of the world $\omega = \{0, 1\}$. Signals are identically distributed and independent, conditional on the true state of the world. There is a common prior $p_0 = \mathbb{P}[\omega = 1] \in (0, 1)$ that the state of the world is 1. Since signals are conditionally independent, all information is contained in the resulting posterior distribution and we will shift attention completely to the latter. Each sender draws his posterior from the probability mass function μ, which is consistent with p_0. We assume that the information structure is such that it leads to a finite number of possible posteriors $\mathcal{P} = \text{supp } \mu$.

For some results, we assume that no signal is uninformative, i.e., $\mu(p_0) = 0$. We call a distribution μ that fulfills this assumption *never-ignorant*. The receiver shares the prior but observes no signal.

In addition to different signals, players are heterogeneous with respect to their preferences, as described by a parameter $\lambda \in \{0, \lambda_0, 1\}$ with $\lambda_0 \in (0, 1)$. Each sender i independently draws a preference parameter λ_i that is independent of the posteriors and distributed according to probability mass function γ. The decision-maker has the commonly known preference parameter λ_0. We call the tuple (p_i, λ_i) the type of sender i, and denote with $\mu \times \gamma$ the distribution over types.

After observing the signal, each sender i simultaneously sends a cheap talk message $t_i \in [0, 1]$ to the receiver. We denote the potentially mixed strategy by $m_i : \mathcal{P} \times \{0, \lambda_0, 1\} \to \Delta[0, 1]$, where $\Delta[0, 1]$ denotes the set of all probability measures over $[0, 1]$. We denote the probability that sender i with type (p_i, λ_i) sends message t_i by $m_i(p_i, \lambda_i)(t_i)$. We call a strategy *truthful* if $m_i(p_i, \lambda_i)(p_i) = 1$ for all types (p_i, λ_i). The tuple of messages of all senders is denoted by $t = (t_1, ..., t_n)$.

The receiver processes the messages of all senders according to Bayes' rule. We denote the belief of the receiver accounting only for sender i's message t_i by $q(t_i)$ and call it the virtual posterior of sender i.[3] The posterior of the receiver incorporating the messages t of all senders is denoted by $q(t)$. After processing all messages, the receiver takes action $a \in \{0, 1\}$. Utilities for senders and the receiver are given by

$$u(a, \omega, \lambda_i) = (1 - \lambda_i)\mathbb{1}\{a = 0\} + \lambda_i \mathbb{1}\{a = 1\} + \mathbb{1}\{a = \omega\},$$

where $\mathbb{1}$ is the indicator function, i.e., $\mathbb{1}\{A\}$ is 1 if event A is true and 0 otherwise.

A player i prefers action 1 if and only if his belief that the state of the world is 1 is larger than or equal to $1 - \lambda_i$. A higher preference parameter λ_i leads to a higher expected utility of player i given that the action is equal to 1. Senders with preference parameters 0 and 1 weakly prefer the action that matches their preference parameter irrespective of the posterior. We call senders with these preference parameters *partisans*. The remaining senders with $\lambda_i = \lambda_0$ have the same interests as the receiver. We call these senders *advisors*.

Before we proceed, we summarize the timing of the game. First, nature draws a state of the world ω. Second, every sender i randomly draws a type (p_i, λ_i) according to the conditional type distribution $\mu_\omega \times \gamma$. Third, each sender i sends a message t_i to the receiver. Last, the receiver takes an action a and payoffs are realized. We assume that the

[3]Different strategies m_i induce different virtual posteriors $q_i(\cdot)$. Anticipating that senders play symmetric strategies in an optimal equilibrium, we drop the subscript i of the virtual posterior $q_i(\cdot)$ to simplify notation.

receiver does not have commitment power, i.e., she can not credibly commit to a decision rule before getting the messages of the senders.[4] Consequently, we solve for perfect Bayesian equilibria.

In the following, we split the analysis into three parts. We start by studying the common interest case in section 3. Here, all senders have aligned preferences. The special case of our setting serves as a benchmark and allows us to get familiar with how the receiver processes the signals from the senders. In section 4, we focus on the information structure of senders, introduce the concept of specialization and illustrate its significance in the common interest case. Lastly, in section 5 we apply these concepts in our analysis of the general case, in which we allow for private interests.

3 Common Interest

In this section, we derive a benchmark equilibrium that maximizes the utility of the receiver when all senders have aligned preferences, i.e., $\gamma(\lambda_0) = 1$ and $\gamma(0) = \gamma(1) = 0$. In the common interest case, such an equilibrium maximizes the utility of the senders, too. The general idea of this equilibrium is straightforward. The receiver needs to perform Bayesian updating given senders' messages, and the senders, knowing that their information will be aggregated in a statistically correct way, can state their posteriors, revealing all their information.

In the description of this equilibrium that follows, we focus on the statistical properties and interpretation of how the receiver updates the information and how she translates it into her decision.

Definition 1. *A receiver follows a weighted majority rule if her strategy $a : [0,1]^n \to \{0,1\}$ is of the form*

$$a(t) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} w(t_i) > \tau \\ 0 & \text{else} \end{cases}$$

for messages $t = (t_1, ..., t_n)$ of senders, a weighting function $w : [0,1] \to \mathbb{R}$, and a threshold τ.

Under a weighted majority rule, the receiver transforms every message t_i into a weight $w(t_i)$ and takes decision 1 if the sum of weighted messages is larger than a threshold τ. One can interpret this as the receiver giving the senders free choice over the weights in the image of w and limiting herself to the application of a simple rule. When the size of the image is equal to 2, this comes down to proposing a decision by qualified majority

[4]In particular, this excludes equilibria of the kind discussed in Gerardi et al. [2009].

voting. We will come back to this analogy in section 5.3. The next proposition translates the above-described equilibrium into this language.

Proposition 1. *The following describes a receiver-optimal perfect Bayesian equilibrium:*

- *Advisors message truthfully, i.e.,* $m_i(p_i, \lambda_0)(t_i) = 1$*, for* $t_i = p_i$*.*
- *The receiver follows a weighted majority rule with weighting function*

$$w(x) = \begin{cases} \ln \frac{x}{1-x} - \ln \frac{p_0}{1-p_0} & \text{if } x \in \mathcal{P} \\ 0 & \text{else} \end{cases}$$

and threshold $\tau = -\left(\ln \frac{\lambda_0}{1-\lambda_0} + \ln \frac{p_0}{1-p_0}\right)$.

Proof. See Appendix A. □

In the optimal equilibrium, senders play the truthful strategy to transmit their posterior to the receiver. The receiver has correct beliefs about this and can deduce from the posteriors the entirety of their information. She then translates it into the optimal decision via Bayesian updating, which takes the form of a weighted majority rule. Hence, there cannot be an equilibrium with higher payoffs for the receiver.[5]

The strategy of the receiver is a generalization of the decision rule derived in Theorem 1 in Nitzan and Paroush [1982]. They study a non-strategic setting with a symmetric prior $p_0 = \frac{1}{2}$. In the proof of Proposition 1, the common prior and the conditional i.i.d. posteriors allow to write the updating process of the receiver as a product formula of posteriors. Applying the logarithm to the equation gives the problem an additive structure. Every posterior can be mapped onto a weight that is the log-likelihood ratio of the posterior $\ln \frac{p_i}{1-p_i}$ minus the log-likelihood ratio of the prior $\ln \frac{p_0}{1-p_0}$. The threshold is the log-likelihood ratio of the receiver's preference parameter $\ln \frac{\lambda_0}{1-\lambda_0}$ minus the log-likelihood ratio of the prior $\ln \frac{p_0}{1-p_0}$. This way, the prior p_0 is taken into account only once (in the threshold) and all other weights are taken as the net of the information from the prior. The log-likelihood ratio of the preference parameter $\ln \frac{\lambda_0}{1-\lambda_0}$ guarantees that action 1 is taken if and only if the final posterior $q(t)$ is larger than $(1 - \lambda_0)$. Thus, the decision rule in the receiver-optimal equilibrium can be interpreted as a weighted majority rule with weighting function $w(x) = \ln \frac{x}{1-x} - \ln \frac{p_0}{1-p_0}$ and threshold $\tau = -\left(\ln \frac{\lambda_0}{1-\lambda_0} + \ln \frac{p_0}{1-p_0}\right)$.

Figure 1 illustrates the weighting function with prior $p_0 = \frac{3}{4}$ for the common interest case. A posterior p_i of sender i that equals the prior p_0 gets weight 0 because it does not transmit any additional information. In contrast, a posterior $p_i \in \{0, 1\}$ means that sender i knows the state of the world perfectly. This sender's information is sufficient to

[5] McLennan [1998] studies optimality of equilibria in common interest games more generally.

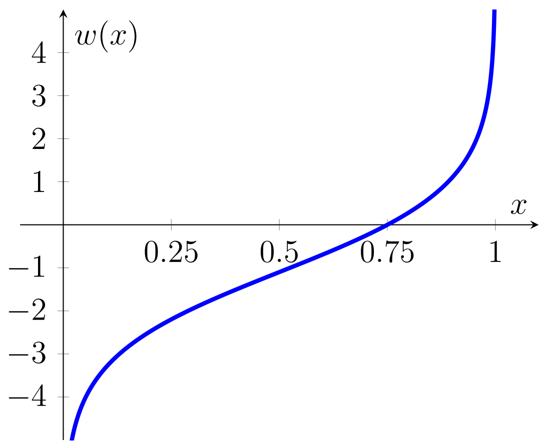


Figure 1: Weighting function $w(x) = \ln \frac{x}{1-x} - \ln \frac{p_0}{1-p_0}$ with prior $p_0 = \frac{3}{4}$ for the common interest case.

make an optimal decision, and he should outweigh all other senders. Thus, as p_i goes to 1 (0), the corresponding weight tends to ∞ ($-\infty$). The unrestrictedly high weight encodes the extraordinary value of perfect information.

In the next section, we refer to the receiver-optimal equilibrium when we assess different distributions of sender types. The expected utility of the receiver $u^\star(q(t))$ with the posterior $q(t)$ is given by

$$u^\star(q) = \begin{cases} \lambda_0 + q & \text{if } q > 1 - \lambda_0 \\ 2 - \lambda_0 - q & \text{else.} \end{cases}$$

We now turn to the analysis of the senders' information structure.

4 Specialization

The posterior distribution of the senders is a crucial object in our model. In this section, we explore its effect on the receiver's ability to match her decision to the state of the world. To this end, we introduce a novel conceptualization of specialization in knowledge as a particular case of Blackwell domination, which we characterize as learning more about the quality of one's signal. To start let us recall this well-known incomplete order.

Definition 2. *Let μ and ν be two posterior distributions with cdfs F and G, respectively. We say that μ is more informative than ν, denoted by $\mu \succ \nu$, if*

$$\int_0^y F(x)\mathrm{d}x \geq \int_0^y G(x)\mathrm{d}x \quad \textit{for all } y \in [0,1]. \tag{1}$$

Blackwell [1950] establishes the concept of informativeness and gives a series of characterizations. The one we use in the above definition is the inverse of the second stochastic order used in decision theory[6]. We will also refer to some of the other equivalent formulations and have gathered them in Appendix B. By the common prior assumption, both distributions have an expected value of p_0.

A second method to compare posterior distributions is through their average informativeness.

Definition 3. *The average informativeness $\pi(\mu)$ of a sender's posterior distribution μ is*

$$\pi(\mu) = \mathbb{E}\left[|p_i - p_0|\right].$$

The average informativeness of a distribution measures the expected distance of the posterior from the prior p_0. Mathematically, it is the first absolute central moment. The greater the difference between prior and posterior, the more precise the information of a sender. A distribution μ with $\pi(\mu) = 0$ does not contain any information at all, whereas the maximal average informativeness is $2p_0(1-p_0)$. Note that higher average informativeness of μ compared to ν is a necessary but not a sufficient condition for μ to be more informative than ν.

Lastly, we combine the two previous concepts in order to define specialization.

Definition 4. *Let μ and ν be two posterior distributions. We say that μ is more specialized than ν, denoted by $\mu \succ_s \nu$, if $\mu \succ \nu$ and $\pi(\mu) = \pi(\nu)$.*

By this definition, a specialist, is more informed not because his posterior is on average further away from the prior, but because his posteriors are more heterogeneous. To illustrate this, let us give a basic example of two senders with symmetric priors. The less specialized sender receives a signal that is accurate 60% of the time in both states, and thus its average informativeness is 0.1. The other sender may receive perfect information 20% of the time and pure noise otherwise. His average informativeness is also 0.1, but his information structure is more informative than the first sender's. The easiest way to see this is to imagine that both receive the same signal, yet the more specialized sender also receives information on which 20% of signals are perfect and 80% are pure noise. We will formalize this idea of specialization as learning about the quality of signals in the following proposition.

Proposition 2. *Let μ and ν be two posterior distributions. Then $\mu \succ_s \nu$ if and only if there exists a coupled measure Π on $[0,1]^2$ with marginals μ and ν s.t. the random posterior vector (p, p') drawn from Π fulfills, the following conditions:*

[6] A "riskier" posterior distribution contains more information.

- *p is a mean-preserving spread of p': $E[p - p'|p'] = 0$ for all p'.*

- *The spread never changes the direction of information: $\mathbb{P}\left(p \in [0, p_0]|(p' \in [0, p_0]\right) = \mathbb{P}\left(p \in [p_0, 1]|(p' \in [p_0, 1]\right) = 1$*

Proof. See Appendix A. □

A more specialized sender receives a signal indicating the wrong state just as often as his less generalized counterpart, yet is better informed about its intensity. For a specialist, judging whether the question at hand is in one of his fields of expertise is very informative. The same is not true for a complete generalist, since the quality of his judgment is uniform across all subjects. Given this characterization, we can now deduce the maximal and minimal elements of the specialization order.

Corollary 1. *We denote with δ_x a Dirac mass at point x. The posterior distribution*

$$\hat{\mu}_\alpha = \frac{\alpha}{2p_0}\delta_0 + \left(1 - \frac{\alpha}{2p_0(1-p_0)}\right)\delta_{p_0} + \frac{\alpha}{2(1-p_0)}\delta_1$$

is more specialized than any other posterior distribution μ with the same average intensity $\pi(\mu) = \alpha$.

Further, the posterior distribution

$$\bar{\mu}_{\alpha,\beta} = \beta\delta_{\frac{2\beta p_0 - \alpha}{\beta}} + (1-\beta)\delta_{\frac{2p_0 - 2\beta p_0 + \alpha}{2-2\beta}}$$

is less specialized than any other posterior distribution μ with average intensity $\pi(\mu) = \alpha$ and $\mu([0, p_0]) = \beta$.

Proof. See Appendix A. □

The most specialized information structures only contain perfect or pure noise signals, while the most generalized ones have exactly one realization indicating each state.

These extreme elements of the specialization order lend themselves to simple interpretations of the ensuing play in the optimal equilibrium under common interest. Under the most specialized measure $\hat{\mu}_\alpha$, the decision is effectively delegated to all senders, with the possibility to abstain[7], and a default option that is optimal given the prior. In contrast, under the least specialized measure, the receiver's optimal decision is equivalent to a qualified majority rule. Specifically, in the optimal equilibrium, the receiver takes action 1 if $\sum_{i=1}^{n} w(t_i) > \tau$ with $w(t_i)$ and τ as in Proposition 1. If the distribution has only two mass points $l \in [0, p_0)$ and $r \in (p_0, 1]$, there are only two weights $\kappa_l = w(p_l)$ and

[7] Note that there will never be two conflicting messages by senders on path.

$\kappa_r = w(p_r)$ in equilibrium. We denote the number of senders whose message is left (right) of the prior by n_l (n_r). The receiver takes action 1 if

$$n_l\kappa_l + n_r\kappa_r > \tau.$$

This corresponds to a qualified majority rule with threshold $n_1(n)$, where n is the total number of senders. The receiver takes action 1 if and only if

$$n_1(n) > \frac{\tau - n\kappa_l}{\kappa_r - \kappa_l}.$$

In the remainder of this subsection, we apply the above insights to obtain the bounds of the receiver's utility. As discussed, the most and least specialized distributions $\hat{\mu}_\alpha$ and $\bar{\mu}_{\alpha,\beta}$ have a particularly simple form. Since more informativeness of individuals' posterior distributions produces more informativeness overall Blackwell and Girshick [1979] (see Proposition C in Appendix B), the utility of the receiver increases with the specialization of posteriors. Thus, the next corollary links corollary 1 with the utility of the receiver.

Corollary 2. *Let μ be a posterior distribution. The utility of the receiver facing senders with posterior distribution μ is bounded above by the utility of a receiver facing senders with $\hat{\mu}_{\pi(\mu)}$, and below by the utility of a receiver facing senders with $\bar{\mu}_{\pi(\mu),\mu[0,p_0]}$.*

We illustrate Proposition 2 in Figure 2. The x-axis represents the number of senders and the y-axis the expected utility of the receiver. The receiver matches the state of the world with probability close to 1 as the number of senders tends to infinity. This holds for any distribution μ with positive average informativeness $\pi(\mu)$. Thus, the blue line (representing the most specialized distribution) and the red line (representing the least specialized distribution) converge to the utility in a setting where the state of the world is known by the receiver. Yet under the most specialized measure, the receiver's utility is higher for any fixed number of senders, illustrating the benefits she reaps from having more specialized senders.

We now turn to the general setting that allows for private interest. We return to a discussion of specialization in the general setting in subsection 5.3.

5 Fake Experts - Private Interest Analysis

In this section, we turn to the case with private interests. In subsection 5.1, we solve for the receiver-optimal equilibrium. In subsection 5.2 we contrast our results with a scenario

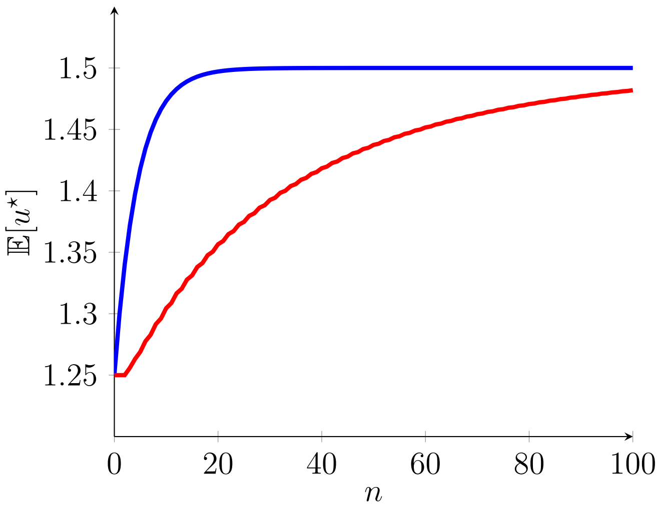


Figure 2: Receiver's utility for n senders given a prior $p_0 = \frac{1}{2}$, average informativeness $\frac{1}{10}$ and $\lambda_0 = \frac{3}{4}$. The blue line represents the utility of the most specialized distribution $\hat{\mu}_{\frac{1}{10}}$ and the red line the least specialized symmetric distribution $\bar{\mu}_{\frac{1}{10},\frac{1}{2}}$.

in which the receiver observes preferences. This allows us to decompose the total loss of information into a loss of signals on the one hand and a despecialization of remaining signals on the other. Lastly, in subsection 5.3, we present two consequences of despecialization. We find that voting is optimal if preferences are sufficiently heterogeneous. Further, average informativeness becomes more and specialization less important as the number of partisans increases.

5.1 Receiver-Optimal Equilibrium

We start by sketching the argument that the strategies in Proposition 1 no longer form an equilibrium in the presence of private interests. Given the strategy of the receiver, partisans maximize the probability that their preferred action is taken by sending the message with the highest possible weight in the respective direction. But then the receiver cannot rationally believe that these messages only come from senders who indeed received the corresponding signals. Consequently, players adapt their strategies. We capture partisans' strategies to choose messages by maximizing their weight in the following definition.

Definition 5. *A sender's strategy, conditional on being partisan, imitates and devalues expertise if a 0- (1-) partisan only mixes over messages with the lowest (highest) virtual posteriors. We denote the resulting lowest (highest) virtual posteriors by* $\underline{b}, \overline{b} \in [0,1]$ *and refer to them as expertise bounds.*

Since the average posterior of a partisan sender still equals the prior, their strategy shifts virtual posteriors towards the prior. It turns out that the best way for the aligned senders to counteract this loss of information is to message truthfully, as summarized in the following theorem.

Theorem 1. *The following describes a receiver-optimal perfect Bayesian equilibrium. There exist unique expertise bounds $\underline{b}, \overline{b} \in [0,1]$, s.t.*

- *Advisors message truthfully, i.e., $m_i(p_i, \lambda_0)(t_i) = 1$, for $t_i = p_i$.*
- *Partisans imitate and devalue expertise:*

$$m_i(p_i,0)(t_i) = \begin{cases} \frac{\gamma(\lambda_0)\mu(t_i)(\underline{b}-t_i)}{\gamma(0)(p_0-\underline{b})} & \text{if } t_i \in \mathcal{P} \wedge t_i \leq \underline{b} \\ 0 & \text{else} \end{cases}$$

$$m_i(p_i,1)(t_i) = \begin{cases} \frac{\gamma(\lambda_0)\mu(t_i)(t_i-\overline{b})}{\gamma(1)(\overline{b}-p_0)} & \text{if } t_i \in \mathcal{P} \wedge t_i \geq \overline{b} \\ 0 & \text{else} \end{cases}$$

- *The receiver uses weighted majority rule with weight function*

$$w(x) = \begin{cases} \ln \frac{\underline{b}}{1-\underline{b}} - \ln \frac{p_0}{1-p_0} & \text{if } x \in \mathcal{P} \wedge x < \underline{b} \\ \ln \frac{x}{1-x} - \ln \frac{p_0}{1-p_0} & \text{if } x \in \mathcal{P} \wedge x \in [\underline{b}, \overline{b}] \\ \ln \frac{\overline{b}}{1-\overline{b}} - \ln \frac{p_0}{1-p_0} & \text{if } x \in \mathcal{P} \wedge x > \overline{b} \\ 0 & \text{else} \end{cases}$$

and threshold $\tau = -\left(\ln \frac{\lambda_0}{1-\lambda_0} + \ln \frac{p_0}{1-p_0}\right)$.

In the equilibrium, advisors play the truthful strategy, as in Proposition 1. It is in their best interest to transmit as much information as possible. Partisans interfere in this communication. Their strategy is independent of their signals. They do not transmit any information to the receiver, but maximize their influence by imitating advisors with the most informative signals. Therefore, the receiver needs to discount these messages. This way, expertise bounds $\underline{b}$ and $\overline{b}$ arise. They constitute bounds on the highest (lowest) possible virtual posteriors associated with any messages. The weights $w(\underline{b})$ and $(w(\overline{b}))$ are the lowest and highest weights used in the weighted majority rule of the receiver. They are endogenously determined by the sender type distribution.

Between the expertise bounds, communication between advisors and the receiver is noise-free, because partisans do not imitate advisors with imprecise signals. Thus, communication is perfect within these bounds, as in the equilibrium from Proposition 1. Off-equilibrium messages receive weight 0. Figure 3 depicts an example of a weighting function of virtual posteriors with upper and lower expertise bounds. The dashed line is the weighting function of the receiver in the absence of partisans.

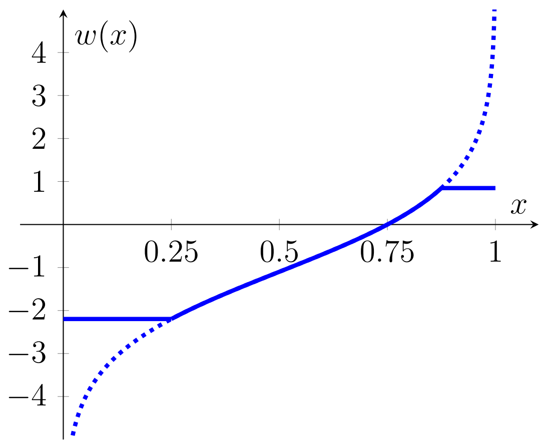


Figure 3: Weighting function $w(x) = \ln \frac{x}{1-x} - \ln \frac{p_0}{1-p_0}$ with prior $p_0 = \frac{3}{4}$, $\underline{b} = \frac{1}{4}$, and $\overline{b} = \frac{7}{8}$.

In the following, we focus on the proof of the optimality of the equilibrium we describe. The proofs regarding the expertise bounds and the fact that the strategies described form an equilibrium are detailed in Appendix A.

Proof of the Optimality of the Equilibrium in Theorem 1. We show that the equilibrium in Theorem 1 is optimal for the receiver. We proceed in two steps. First, we introduce a technique that allows us to compare equilibria in the common interest case. We could have used this technique to prove Proposition 1, but found the proof above more constructive. Second, we show that the comparison carries over to the case with partisans. More concretely, we show that if an equilibrium in which advisors play the truthful strategy is more informative than another one in the case without partisans, it continues to be more informative than the other one in the presence of partisans.

The receiver bases her decision on the virtual posteriors $q(t_i)$ which she infers from messages t_i of senders $i = \{1, ..., n\}$. The same set of virtual posteriors leads to the same decision. The distribution of virtual posteriors $q(t_i)$ for sender i is determined by the distribution of posteriors μ and the sender i's strategy m_i.

Definition 6. *Let μ be a distribution of posteriors and m_i the strategy of sender i. We denote the distribution of virtual posteriors of sender i by $\mu^{\gamma}_{m_i}$ and define it by its cdf*

$$F^{\gamma}_{m_i}(x) = \mathbb{P}[q(t_i) \leq x]$$

where t_i is sender i's message. We suppress superscript γ in the common interest case, i.e., we write μ_{m_i} and F_{m_i} if $\gamma(\lambda_0) = 1$.

In the following, we compare the virtual posterior distribution of the equilibrium in which advisors play the truthful strategy with virtual posterior distributions of other equilibria. We know from Proposition 1 that playing the truthful strategy is part of a receiver-optimal equilibrium for the common interest case. Using the concept of virtual posterior distributions helps us generalize this observation to the case with partisans.

We start with the analysis of the common interest case. Note that other equilibria lead to the same distribution of virtual posteriors as in Proposition 1. Consequently, these equilibria induce the same ex-ante expected utility of the receiver. For example, consider strategies $\tilde{m}_i$ characterized by $\tilde{m}_i(p_i, \lambda_i) = 1 - m_i(p_i, \lambda_i)$ for senders $i = \{1, ..., n\}$ and a receiver strategy in which she uses a weighting function $\tilde{w}$ characterized by $\tilde{w}(x) = w(-x)$. We select the equilibrium in which advisors play the truthful strategy as representative of all equilibria that induce this distribution of virtual posteriors.

Lemma 1. *Let μ be a posterior distribution, $m_i^\star$ the truthful strategy, and m_i' any other strategy. Then it holds that $\mu_{m_i^\star}$ is more informative than $\mu_{m_i'}$, i.e., $\mu_{m_i^\star} \succ \mu_{m_i'}$.*

Proof. See Appendix A. □

The proof of Lemma 1 builds on the machinery of Blackwell and Girshick [1979]. If $m_i^\star$ is the truthful strategy, any distribution $\mu_{m_i'}$ that is induced by another strategy m_i' can be constructed from the distribution $\mu_{m_i^\star}$ by an application of garblings. The only way in which the virtual posterior distribution $\mu_{m_i'}$ can differ from $\mu_{m_i^\star}$ is that a sender i might send a message t_i for two different posteriors p_i and p_i'. The virtual posterior $p(t_i)$ is a weighted average of posteriors that induce sending t_i. Hence, strategy $m_i^\star$ is a garbling of $m_i^\star$ which implies that $\mu_{m_i'}$ is a garbling of $\mu_{m_i^\star}$. Under Theorem 12.3.2 in Blackwell and Girshick [1979] (see Proposition B in Appendix B), the sender-wise comparison is sufficient to permit the conclusion that the posterior distribution $q(t)$ in the equilibrium in which senders play the truthful strategy is more informative than that in any other equilibrium. Proposition 1 follows by Theorem 12.2.2 (4) in Blackwell and Girshick [1979] (see Proposition C in Appendix B). More informative distributions imply higher ex-ante expected utility for the receiver.

In the last step, we show that the argument can be generalized to the case with partisans. In particular, this implies that the virtual posterior distribution $\mu^\gamma_{m_i^\star}$ is more informative than any other distribution $\mu^\gamma_{m_i'}$.

Lemma 2. *Let μ be a posterior distribution, $m_i^\star$ a strategy in which advisors play truthfully, and m_i' any other strategy. Then, if $\mu_{m_i^\star}$ is more informative than $\mu_{m_i'}$, it follows that $\mu^\gamma_{m_i^\star}$ is more informative than $\mu^\gamma_{m_i'}$, i.e.*

$$\mu_{m_i^\star} \succ \mu_{m_i'} \Rightarrow \mu^\gamma_{m_i^\star} \succ \mu^\gamma_{m_i'}.$$

Proof. See Appendix A. □

No matter the strategy of advisors, partisans imitate and devalue expertise. This effect can only be diminished if advisors have already garbled the most valuable signals themselves, but in sum, this can never increase the informativeness of the virtual posterior distribution.

We have shown that the virtual posterior of sender i is most informative if types with $\lambda_i = \lambda_0$ play the truthful strategy. Again, under Theorem 12.3.2 in Blackwell and Girshick [1979] (see Proposition B in Appendix B), the sender-wise comparison carries over to the overall information structure. By Theorem 12.2.2 (4) in Blackwell and Girshick [1979] (see Proposition C in Appendix B), we conclude that there cannot be any better equilibrium for the receiver than the equilibrium described in Theorem 1. This concludes the proof.

□

5.2 Lack of Trust versus Lack of Competence

In this subsection, we contrast the equilibrium we just described to the case in which the preferences of senders are known to the receiver. In this latter case, partisans cannot transmit any information via cheap talk, since their preferences are known to be state-independent. As in the case with unknown preferences, their signal is lost to the receiver. From her perspective, it is as if partisans are incompetent, i.e., never receive a signal. However, the situation is still better than that under unknown preferences, in which partisans exert influence by imitating and thereby undermining trust in aligned senders with the most informative signals. In that sense, lack of trust in an expert is worse than his potential lack of competence.

Theorem 2. *Let $\mu^\gamma_{unknown}$ and μ^γ_{known} be the virtual posterior distribution in the equilibrium above and the most informative equilibrium with known preferences respectively. The loss of average informativeness in both situations is identical and equal to the share of partisan senders:*

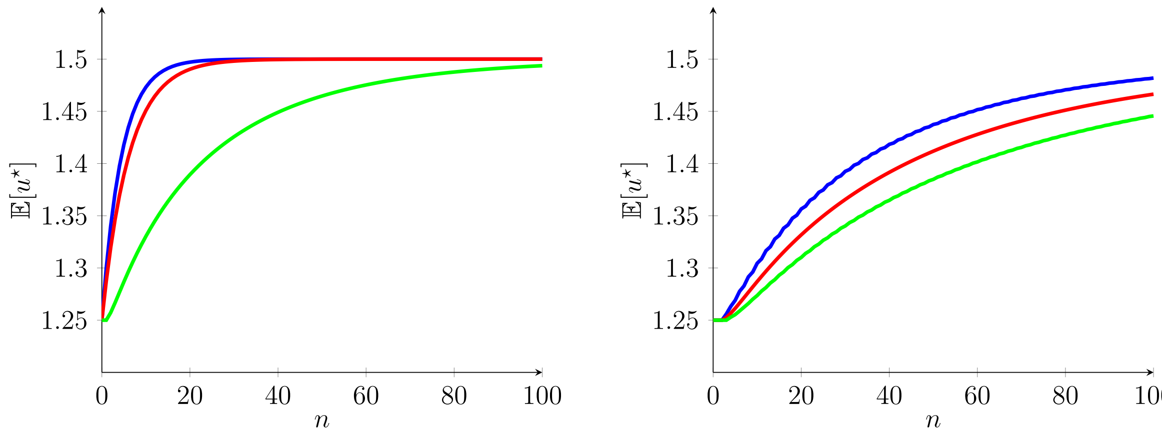
$$\pi(\mu^\gamma_{unknown}) = \pi(\mu^\gamma_{known}) = \gamma(\lambda_0)\pi(\mu)$$

Yet the virtual posteriors under unknown partisanship are less informative than under known partisanship:

$$\mu^\gamma_{unknown} \prec_s \mu^\gamma_{known}$$

Proof. See Appendix A □

The result allows us to decompose the effect of unknown partisanship into a loss of signals and a loss in specialization. While the first phenomenon affects all posterior distributions in very similar ways, the effect of despecialization is strongest for senders with very specialized posterior measures. We have illustrated this in Figure 4.



(a) Most specialized senders: $\hat{\mu}_{\frac{1}{10}}$

(b) Least specialized symmetric senders: $\bar{\mu}_{\frac{1}{10},\frac{1}{2}}$

Figure 4: Receiver's utility for n senders given prior $p_0 = \frac{1}{2}$, average informativeness $\frac{1}{10}$ and $\lambda_0 = \frac{3}{4}$. The blue lines represent the common interest benchmark. The red and green lines represent the case with $\gamma(0) = \gamma(1) = \frac{1}{8}$, where partisanship is known and unknown to the receiver, respectively.

Having studied the two distinct parts of information loss, we continue by pointing out two rather stark consequences of high degrees of partisanship. We demonstrate that sufficiently heterogeneous preferences can prevent any differentiating weighting of messages. Further, we show that the value of specialization vanishes and average informativeness becomes more important as the share of partisans rises.

5.3 A Justification for Voting

In this subsection, we illustrate the effect of private interests on information aggregation and the utility of the receiver. We begin with a result that states that a sufficient number of partisans can prevent transmission of any information that is finer than the mere direction of the preferred alternative. In other words, all senders send one of only two messages in the optimal equilibrium. This small message space can be interpreted as voting.

Proposition 3. *Let μ be never-ignorant. Then there exists $c_0, c_1 \in (0,1)$ with $c_0 + c_1 < 1$, s.t. for all γ with $\gamma(0) \geq c_0$ and $\gamma(1) \geq c_1$ the receiver forms only two expected posteriors, i.e., voting is the most informative equilibrium.*

Proof. See Appendix A. □

The proof of Proposition 3 exploits properties of the expertise bounds. We find c_0 and (c_1) such that the lower (upper) expertise bound coincides with the highest (lowest) possible posterior that is smaller (greater) than the prior. This means that all posteriors on one side of the prior get the same weight. To guarantee that $c_0 + c_1 < 1$, we need to assume that μ is never-ignorant. Note that if $c_0 + c_1 = 1$ all posteriors get the same weight $w(p_0)$, because all senders are partisans who do not send informative messages.

Figure 5 illustrates the effect of different levels of partisans. Panel (a) and (b) depict the case without partisans. A 1% share partisans (Panels [c] and [d]) devalues the weight of advisors with the most precise posteriors. A 5% share of partisans (Panels [e] and [f]) prevent the differentiation of advisors with the two most precise posteriors on both sides of the prior. Any differentiation between advisors of one side of the prior breaks down if the share of partisans is 25% (Panels [g] and [h]) or more. Then, only the direction of the posterior can be transmitted. This case corresponds to the situation in Proposition 3.

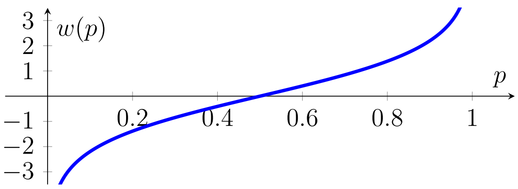
The next result points out that average informativeness becomes more important as the share of partisans increases and the effects of heterogeneously-distributed posteriors decreases. Concretely, a posterior distribution with higher average informativeness leads to higher receiver utility if the share of partisans is sufficiently high.

Proposition 4. *Let μ and ν with $\pi(\mu) > \pi(\nu)$ be never-ignorant posterior distributions with cdfs F and G, respectively. Then there exist $c_0, c_1 \in (0, 1)$ with $c_0 + c_1 < 1$, s.t. for all γ with $\gamma(0) \geq c_0$ and $\gamma(1) \geq c_1$ and any number of senders n the ex-ante expected utility of the sender is greater under posterior distribution μ than under ν.*

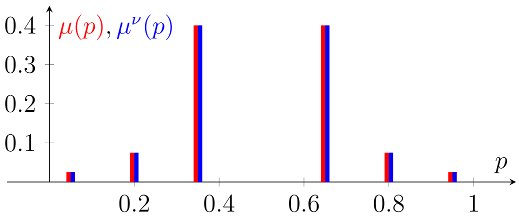
Proof. See Appendix A. □

The proof of Proposition 4 builds on Proposition 3. Suppose there are sufficiently many partisans so that voting is the optimal equilibrium for both distributions. The virtual posterior of any message of senders with the higher average informativeness is more precise, i.e., further away from the prior p_0. This implies that the ex-ante utility of the sender is higher under the posterior distribution with higher average informativeness.

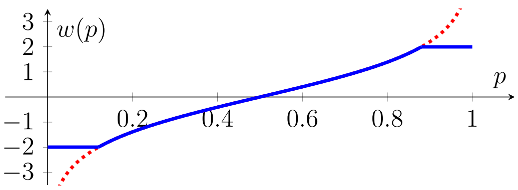
Proposition 4 contrasts the observation on the value of specialization in the common interest case in section 4. Without partisans, experts with more heterogeneously distributed posteriors but lower average informativeness can be better for the receiver. This statement is not true if the share of partisans increases sufficiently. Then, higher average informativeness is all that matters and all value from specialization is lost. We illustrate this effect in Figure 6.



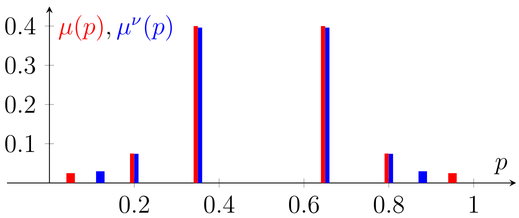
(a) Weighting function with $\gamma(0) = \gamma(1) = 0$



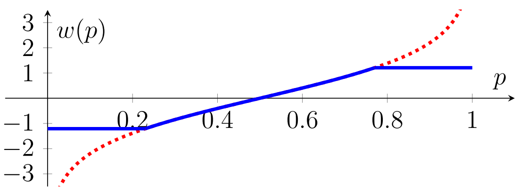
(b) Virtual pmf with $\gamma(0) = \gamma(1) = 0$



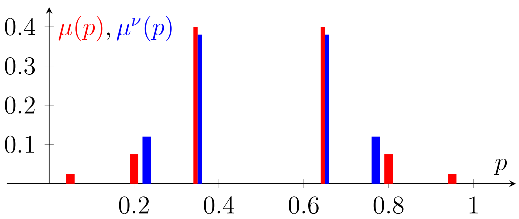
(c) Weighting function with $\gamma(0) = \gamma(1) = 0.005$



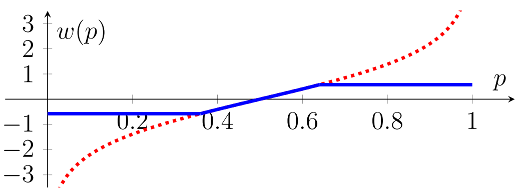
(d) Virtual pmf with $\gamma(0) = \gamma(1) = 0.005$



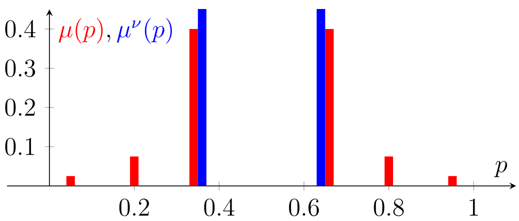
(e) Weighting function with $\gamma(0) = \gamma(1) = 0.025$



(f) Virtual pmf with $\gamma(0) = \gamma(1) = 0.025$

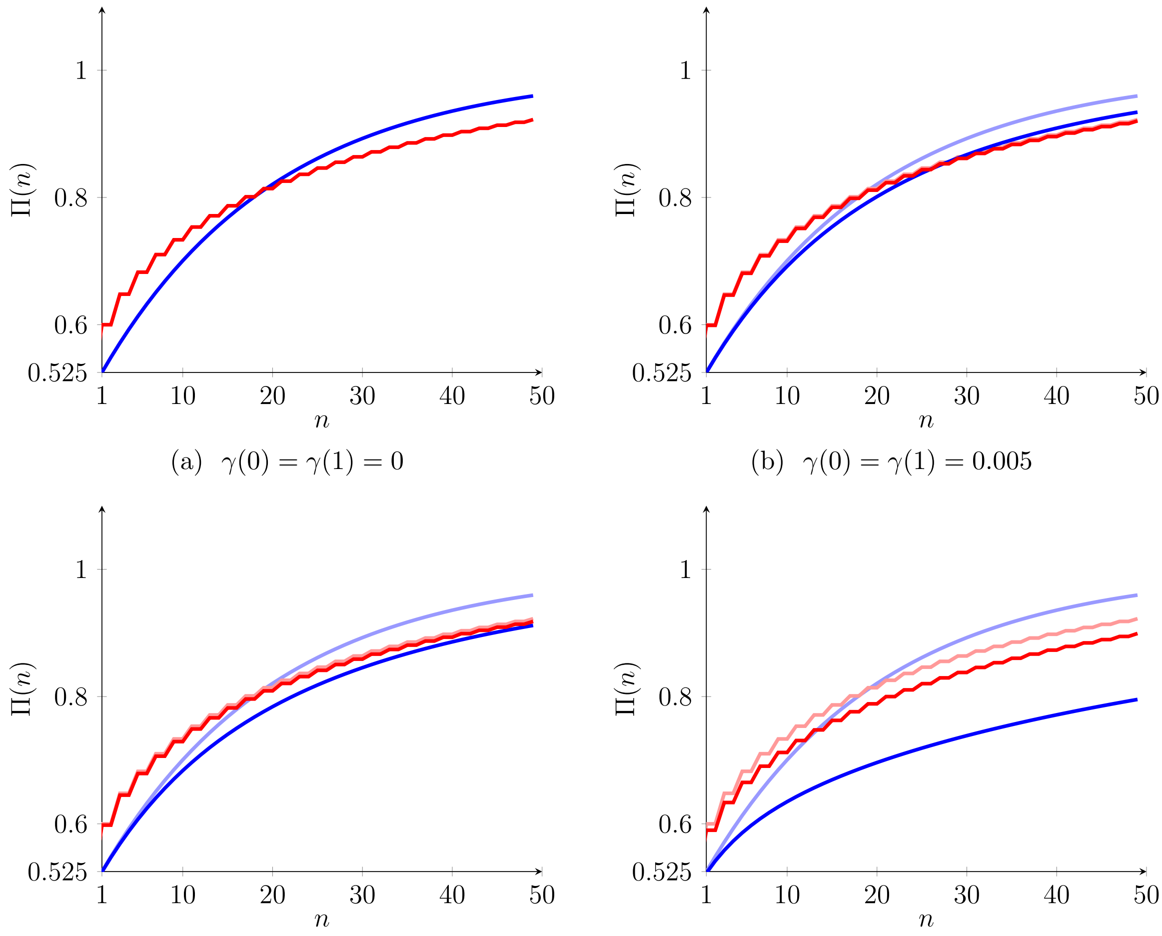


(g) Weighting function with $\gamma(0) = \gamma(1) = 0.125$



(h) Virtual pmf with $\gamma(0) = \gamma(1) = 0.125$

Figure 5: Equilibrium weighting function and virtual probability mass functions for probability mass function μ with $\mathbb{P}\left[p = \frac{1}{20}\right] = \mathbb{P}\left[p = \frac{19}{20}\right] = \frac{1}{40}$, $\mathbb{P}\left[p = \frac{1}{5}\right] = \mathbb{P}\left[p = \frac{4}{5}\right] = \frac{3}{40}$, and $\mathbb{P}\left[p = \frac{7}{20}\right] = \mathbb{P}\left[p = \frac{13}{20}\right] = \frac{4}{10}$. This pmf is symmetric around prior $p_0 = \frac{1}{2}$.



(a) $\gamma(0) = \gamma(1) = 0$ (b) $\gamma(0) = \gamma(1) = 0.005$

(c) $\gamma(0) = \gamma(1) = 0.01$ (d) $\gamma(0) = \gamma(1) = 0.05$

Figure 6: The blue lines correspond to a pmf μ_F with $\mathbb{P}[p=0] = \mathbb{P}[p=1] = 0.025$, and $\mathbb{P}\left[p=\frac{1}{5}\right] = \mathbb{P}\left[p=\frac{1}{2}\right] = 0.95$. The red lines correspond to a pmf μ_G with $\mathbb{P}\left[p=\frac{2}{5}\right] = \mathbb{P}\left[p=\frac{3}{5}\right] = \frac{1}{2}$. Both pmfs imply that both states of the world are equally likely, i.e., $p_0 = \frac{1}{2}$.

5.4 Asymmetric Senders

In the previous discussion, we have focused on the scenario in which all senders are ex-ante symmetric. However, in some of the applications, it stands to reason that the receiver can discriminate between the senders based on prior knowledge. A regulator might understand that one of his advising doctors has previously worked on the approval of similar drugs and might hence believe his posterior distribution to be more informative than average. Similarly, a judge might rightly assume that a relative of the defendant is more likely to be biased in the defendant's favor than against it.

As we have seen in our discussion of the symmetric senders, the receiver's learning from messages happens sender by sender, i.e., the message of one sender does not change how the message of another is processed. Consequently, all that changes in our model, when we allow for asymmetric senders, is the weighting function that the receiver uses needs to be individualized based on each sender's posterior and preference distribution.

Learning about the informativeness and preferences of senders can of course only improve the receiver's situation, since she always can choose to ignore that information regardless. Hence, our symmetric case can also be interpreted as a worst-case benchmark for the effect of partisanship on information transmission.

Departing from that benchmark, our previous analysis suggests that the largest gains from knowledge about individual sender's informativeness and preferences are generated by the possibility of finding a specialist with a low probability of partisanship. As we have seen, specialization can have great benefits as long as the expert is also well trusted.

6 Conclusion

Our ambition in this research has been to understand the optimal communication of a decision-maker with multiple advising experts when she faces uncertainty about experts' preferences. In particular, we have been interested in how these uncertainties effect changes between senders with different information structures.

We have found that communication that discriminates between messages, indicating different degrees of confidence, is potentially very informative for the receiver, but also highly vulnerable to strategic manipulation by partisan experts. Consequently, such communication is not optimal in a case with high levels of partisanship. In contrast, binary communication protocols such as voting prove to be very robust, explaining their prevalence as a means for information aggregation.

Our research has also opened the gate for questions regarding political lobbying. Much of the literature on the subject (see for example Buchanan et al. [1980] and Baye et al. [1993]) has focused on lobbying as a way in which special interest groups try to provide incentives for political actors, in order to sway them in their favored direction. It is, however, just as plausible for such groups to buy influence with advising experts to influence politicians' beliefs rather than offer direct incentives. Our work shows that this can be effective even if politicians are aware of it, as long as they remain ignorant about the exact identity of the experts who have been compromised. In particular, interest groups may seek to sometimes influence experts against their own favored decision to create the justified belief that some experts advocating the other side are not trustworthy. When talk is cheap, trust is s valuable yet vulnerable asset.

A Appendix: Proofs

Proof of Proposition 1. As noted in the text, in the following we use the proof technique of Theorem 1 in Nitzan and Paroush [1982], who derive the optimal non-strategic processing of signals with a symmetric prior $\lambda_0 = \frac{1}{2}$.

In the main text, we use the same notation for random variables and their realizations. For this proof, it is useful to introduce a separate notation. We use upper-case characters for random variables and lower-case characters for their realizations.

The receiver processes messages t to update her posterior. She prefers the action that yields the higher expected utility given her posterior $q(t)$. More precisely, an optimal decision rule selects action 1 if

$$
\begin{aligned}
& & \lambda_0 + \mathbb{P}[\omega = 1|T = t] &> (1-\lambda_0) + \mathbb{P}[\omega = 0|T = t] \\
&\Leftrightarrow & \lambda_0 \mathbb{P}[\omega = 1|T = t] &> (1-\lambda_0)\mathbb{P}[\omega = 0|T = t] \\
&\Leftrightarrow & \lambda_0 \frac{\mathbb{P}[P = p|\omega = 1] \cdot \mathbb{P}[\omega = 1]}{\mathbb{P}[P = p]} &> (1-\lambda_0)\frac{\mathbb{P}[P = p|\omega = 0] \cdot \mathbb{P}[\omega = 0]}{\mathbb{P}[P = p]} \\
&\Leftrightarrow & \lambda_0 p_0 \prod_i \mathbb{P}[P_i = p_i|\omega = 1] &> (1-\lambda_0)(1-p_0)\prod_i \mathbb{P}[P_i = p_i|\omega = 0] \\
&\Leftrightarrow & \lambda_0 p_0 \prod_i \frac{p_i}{p_0} &> (1-\lambda_0)(1-p_0)\prod_i \frac{1-p_i}{1-p_0} \\
&\Leftrightarrow & \prod_i \left(\frac{p_i}{1-p_i}\frac{1-p_0}{p_0}\right) &> \frac{1-\lambda_0}{\lambda_0}\frac{1-p_0}{p_0} \\
&\Leftrightarrow & \sum_i \left(\ln\frac{p_i}{1-p_i} - \ln\frac{p_0}{1-p_0}\right) &> -\left(\ln\frac{\lambda_0}{1-\lambda_0} + \ln\frac{p_0}{1-p_0}\right).
\end{aligned}
$$

The first equivalence is a simple algebraic consequence of the fact that the first factors and the second factors each add to one. For the second equivalence, we apply Bayes' rule and exploit the fact that senders play the truthful strategy. In the third step, we use the conditional independence of signals. We arrive at the fifth equation by applying Bayes' rule once again. The sixth equation is a simple reformulation of the fourth. Finally, we obtain the last equation by taking the logarithm on both sides. The resulting decision rule can be interpreted as a weighted majority rule with weighting function

$$
w(t_i) = \begin{cases} \ln\frac{t_i}{1-t_i} - \ln\frac{p_0}{1-p_0} & \text{if } t_i \in \mathcal{P} \\ 0 & \text{else,} \end{cases}
$$

and threshold $\tau = -\left(\ln\frac{\lambda_0}{1-\lambda_0} + \ln\frac{p_0}{1-p_0}\right)$.

It is optimal for senders to play the truthful strategy since senders and the receiver have the same utility function. With the truthful strategy, senders can transmit all available information. Any beneficial transformation of messages can be done by the receiver. □

Lemma 3. *Let $c \in \mathbb{R}$ be a constant and Z be a random variable with $E[Z] = 0$. Then*

$$E[|c + Z|] \geq |c|.$$

and further equality holds if and only if $P(c + Z \geq 0) = 1$ for $c \geq 0$ and $P(c + Z \leq 0) = 1$ for $c \leq 0$.

Proof of Lemma 3. Suppose $c \geq 0$. We then have:

$$\begin{aligned}
E[|c + Z|] =& P(c + Z > 0)E[c + Z|c + Z > 0] + P(c + Z \leq 0)E[-c - Z|c + Z \leq 0] \\
=& P(c + Z \geq 0)E[c + Z|c + Z \geq 0] - P(c + Z \leq 0)E[-c - Z|c + Z \leq 0] \\
& + 2P(c + Z \leq 0)E[-c - Z|c + Z \leq 0] \\
=& c + 2P(c + Z \leq 0)E[-c - Z|c + Z \leq 0]
\end{aligned}$$

The last term equals $|c|$ if and only if $P(c + Z \leq 0)$ and is otherwise strictly greater. The argument for $c < 0$ is symmetrical. □

Proof of Proposition 2. We rely heavily on a similar characterization by Rothschild and Stiglitz [1970], which the reader can find in Appendix B.

From this statement, we can directly deduce:

Let μ and ν be two posterior distributions. Then $\mu \succ \nu$ if and only if there exists a coupled measure Π on $[0, 1]^2$ with marginals μ and ν s.t. the random posterior vector (p, p') drawn from Π fulfills: p is a mean-preserving spread of p': $E[p - p'|p'] = 0$ for all p'.

Hence, it remains to be shown that provided $\mu \succ \nu$, we have $\pi(\mu) = \pi(\nu)$ if and only if the corresponding mean-preserving spread preserves the direction of information: $\mathbb{P}\left(p \in [0, p_0]|(p' \in [0, p_0]\right) = \mathbb{P}\left(p \in [p_0, 1]|(p' \in [p_0, 1]\right) = 1$.

$$\begin{aligned}
\pi(\mu) =& E_p[|p - p_0|] = E_p[|p - p' + p' - p_0|] = E_p[E_{p'}[|p - p' + p' - p_0| \; |p']] \\
\geq& E_{p'}[|p' - p_0|] = \pi(\mu)
\end{aligned}$$

The inequality follows from Lemma 3, and by the same lemma, holds with equality if and only if the mean-preserving spread preserves the direction of information.

□

Proof of Corollary 1. We start with our claim about $\hat{\mu}_\alpha$. Under Proposition 2, any posterior measure with mass in the interval $(0, p_0)$ or $(p_0, 1)$ gets more specialized when this mass is spread, while preserving the mean on $\{0, p_0\}$ and $\{p_0, 1\}$ respectively. The resulting posterior distribution necessarily fulfills the following three conditions:

1. Unit-mass: $\hat{\mu}(0) + \hat{\mu}(p_0) + \hat{\mu}(1) = 1$
2. Identical prior: $0 \cdot \hat{\mu}(0) + p_0 \cdot \hat{\mu}(p_0) + 1 \cdot \hat{\mu}(1) = p_0$
3. Identical average informativeness: $p_0 \cdot \hat{\mu}(0) + (1 - p_0) \cdot \hat{\mu}(1) = \alpha$

These three conditions characterize $\hat{\mu}_\alpha$, as stated.

Let us now analyze $\bar{\mu}_{\alpha,\beta}$. Note first that any posterior measure μ is less specialized than a posterior measure that is constructed by contracting some or all mass from one of the intervals $[0, p_0]$ and $[p_0, 1]$ to the mean of that mass. This is simply the inverse operation of the mean-preserving spreads that are restricted to one side. One can iterate this process as long as there are more than two mass points in either of $[0, p_0]$ or $[p_0, 1]$. The positions of the mass-points for the least specialized measure are determined by the mass allocated to these two intervals and the average intensity. Denote by l and r the position of the left and right mass-points. We then get the following four equations that characterize $\hat{\mu}_{\alpha,\beta}$, as stated.

1. Unit-mass: $\bar{\mu}(l) + \bar{\mu}(r) = 1$
2. Identical prior: $p_l \cdot \bar{\mu}(l) + p_r \cdot \bar{\mu}(r) = p_0$
3. Identical average informativeness: $(p_0 - p_l) \cdot \bar{\mu}(l) + (p_r - p_0) \cdot \bar{\mu}(r) = \alpha$
4. Ratio of 0-favoring signals: $\bar{\mu}(l) = \beta$

□

Lemma 4. *For posterior distribution μ with cdf F and preference distribution γ, the lower expertise bound $\underline{b}$ in the receiver-optimal equilibrium is determined by*

$$\gamma(0)\,(p_0 - \underline{b}) = \int_0^{\underline{b}} (\underline{b} - x)\mathrm{d}\mu = \gamma(\lambda_0) \cdot \int_0^{\underline{b}} F(x)\ \mathrm{d}x,$$

and the upper expertise bound $\overline{b}$ is determined by

$$\gamma(1)\left(\overline{b} - p_0\right) = \int_{\overline{b}}^{!} (X - \overline{b})\mathrm{d}\mu = \gamma(\lambda_0) \cdot \int_{\overline{b}}^{1} 1 - F(x)\ \mathrm{d}x.$$

Proof. To guarantee that the mixing probabilities as laid out in the strategy for 0-partisans in Theorem 1 are probabilities, it must be the case that

$$\frac{\gamma(\lambda_0)\mu(t_i)(\underline{b}-t_i)}{\gamma(0)(p_0-\underline{b})} \geq 0 \quad \text{for all } t_i \leq \underline{b} \wedge t_i \in \mathcal{P}, \text{ and}$$

$$\sum_{t_i \leq \underline{b} \wedge t_i \in \mathcal{P}} \frac{\gamma(\lambda_0)\mu(t_i)(\underline{b}-t_i)}{\gamma(0)(p_0-\underline{b})} = 1.$$

The first condition is fulfilled for all $t_i \leq \underline{b}$. The second condition implies that

$$\begin{aligned}
\gamma(0)(p_0-\underline{b}) =& \gamma(\lambda_0)\int_0^{\underline{b}} (\underline{b}-x)\mathrm{d}\mu \\
=& \gamma(\lambda_0)\cdot F(\underline{b})\cdot \underline{b} - \gamma(\lambda_0)\cdot \left(F(\underline{b})\cdot \underline{b} - \int_0^{\underline{b}} F(x)\mathrm{d}x\right) \\
=& \gamma(\lambda_0)\cdot \int_0^{\underline{b}} F(x)\mathrm{d}x,
\end{aligned}$$

where we obtain the second equation by integration by parts.

The calculation for the upper expertise bound is analogous. □

Proof of existence and uniqueness of expertise bounds in Theorem 1. With Lemma 4 the characterization of the lower expertise bound $\underline{b}$ is given by

$$\gamma(0)\,(p_0-\underline{b}) = \gamma(\lambda_0)\cdot \int_0^{\underline{b}} F(x)\ \mathrm{d}x.$$

Note that the left side of the equation is strictly decreasing in $\underline{b} \in [0, p_0]$ and is 0 only if $\underline{b} = p_0$. The right side is weakly increasing in $\underline{b}$ and is 0 for $\underline{b} = 0$. Further, both sides are continuous in $\underline{b}$. Thus, there is a unique $\underline{b}$ that fulfills the equation.

The proof for the upper expertise bound is analogous.

□

Proof of the equilibrium in Theorem 1. As in the proof of Proposition 1, we use upper-case characters for random variables and lower-case characters for their realizations.

We start to calculate the virtual posterior $q(t_i)$ of the receiver after receiving message t_i. The only senders that send messages within the expertise bounds are advisors. Thus $q(t_i) = t_i$ for $t_i \in (\underline{b}, \bar{b}) \cap \mathcal{P}$. For messages $t_i \leq \underline{b}$ with $t_i \in \mathcal{P}$ the virtual posterior of the receiver is

$$\begin{aligned}
q(t_i) =& \mathbb{P}\left[\omega = 1 | T_i = t_i\right] \\
=& \frac{\sum_{\lambda \in \{0,\lambda_0\}} \mathbb{P}\left[T_i = t_i | \omega = 1 \wedge \Lambda = \lambda\right] \mathbb{P}[\omega = 1]\mathbb{P}\left[\Lambda = \lambda\right]}{\mathbb{P}\left[T_i = t_i\right]}
\end{aligned}$$

$$= \frac{\gamma(\lambda_0)\mu(t_i)t_i + \gamma(0)\frac{\gamma(\lambda_0)\mu(t_i)(\underline{b}-t_i)}{\gamma(0)(p_0-\underline{b})}p_0}{\gamma(\lambda_0)\mu(t_i) + \gamma(0)\frac{\gamma(\lambda_0)\mu(t_i)(\underline{b}-t_i)}{\gamma(0)(p_0-\underline{b})}}$$

$$= \underline{b}.$$

The calculation for the virtual posterior of messages $t_i \leq \underline{b}$ with $t_i \in \mathcal{P}$ is $q(t_i) = \bar{b}$ by an analogous calculation. Thus, the receiver's on-equilibrium beliefs are consistent with Bayes updating.

The technique of Nitzan and Paroush [1982] and the proof of Proposition 1 teach us how to process a set of (virtual) posteriors optimally. Again, the best response of the receiver can be interpreted as a weighted majority rule with weighting function

$$w(x) = \ln\frac{q(x)}{1-q(x)} - \ln\frac{p_0}{1-p_0} = \begin{cases} \ln\frac{\underline{b}}{1-\underline{b}} - \ln\frac{p_0}{1-p_0} & x \in \mathcal{P} \wedge x \leq \underline{b} \\ \ln\frac{x}{1-x} - \ln\frac{p_0}{1-p_0} & x \in \mathcal{P} \wedge \underline{b} \leq x \leq \bar{b} \\ \ln\frac{\bar{b}}{1-\bar{b}} - \ln\frac{p_0}{1-p_0} & x \in \mathcal{P} \wedge \bar{b} \leq x \\ 0 & \text{else} \end{cases}$$

and threshold $\tau = -\left(\ln\frac{\lambda_0}{1-\lambda_0} + \ln\frac{p_0}{1-p_0}\right)$.

We proceed by proving that senders play best responses. Partisans maximize the probability that the receiver will take the action that matches their preference parameter. Given the strategy of advisors and the receiver, they send a message with maximal weight in the preferred direction. In the equilibrium strategy 0-(1-)partisans mix over messages with weight $\ln\frac{\bar{b}}{1-\bar{b}} - \ln\frac{p_0}{1-p_0}$ ($\ln\frac{\underline{b}}{1-\underline{b}} - \ln\frac{p_0}{1-p_0}$) which is the highest (lowest) weight assigned by the receiver. Hence, these partisans play best responses.

We proceed to analyze the best responses of advisors. Suppose an advisor is pivotal, i.e., two different messages of his induce different actions by the receiver. The advisor and the receiver have the same utility function and prefer the same action when they have the same information. Thus, the best the advisor can do is to reveal all his information to the receiver, who processes it optimally. If the advisory sender is not pivotal, any message is a best response.

Taken together, the strategies and updating in Theorem 1 are a perfect Bayesian equilibrium.

□

Proof of Lemma 1. Let $\mu_{m_i^\star}$ be the virtual posterior distribution under the truthful strategy $m_i^\star$. Any distribution $\mu_{m_i'}$ that is induced by another strategy m_i' can be constructed from $\mu_{m_i^\star}$ by an application of garblings. We do not restrict strategies to use only a finite set of messages. Therefore, we apply a result from Blackwell [1953] that generalizes Theorem 12.2.2 in Blackwell and Girshick [1979] (see Proposition A in Appendix B) to the case with continuous signals. Thereby, we conclude that $\mu_{m_i^\star}$ is more informative than $\mu_{m_i'}$. □

Lemma 5. *Let μ and ν with $\mu \succ \nu$ be posterior distributions with cdfs F and G, respectively. Let γ be the distribution of preference parameters. Then, the lower (upper) expertise bound $\underline{b}_\mu$ of μ is weakly smaller (greater) or equal than the lower (upper) expertise bound $\underline{b}_\nu$ of ν in the optimal equilibria with partisans.*

Proof of Lemma 5. Suppose that $\underline{b}_\nu < \underline{b}_\mu$, and use Lemma 4 to see that

$$\begin{aligned}
\gamma(0)\left(p_0 - \underline{b}_\nu\right) &= \gamma(\lambda_0) \cdot \int_0^{\underline{b}_\nu} G(x)\mathrm{d}x \\
&\leq \gamma(\lambda_0) \cdot \int_0^{\underline{b}_\nu} F(x)\mathrm{d}x \\
&\leq \gamma(\lambda_0) \cdot \int_0^{\underline{b}_\mu} F(x)\mathrm{d}x = \gamma(0)\left(p_0 - \underline{b}_\mu\right).
\end{aligned}$$

Hence $\underline{b}_\mu \leq \underline{b}_\nu$, which is a contradiction. The proof for the upper expertise bound is analogous. □

Proof of Lemma 2. To simplify notation, we denote $\mu_{m_i^\star}$ by μ, $\mu_{m_i'}$ by ν, $\mu^\gamma_{m_i^\star}$ by μ^γ, and $\mu^\gamma_{m_i'}$ by ν^γ. Further, we denote $F_{m_i^\star}$ by F, $F_{m_i'}$ by G, $F^\gamma_{m_i^\star}$ by F^γ, $F^\gamma_{m_i'}$ by G^γ.

To prove that μ^γ is more informative than ν^γ, we show that

$$\int_0^y G^\gamma(x)\mathrm{d}x \leq \int_0^y F^\gamma(x)\mathrm{d}x \qquad \text{for all } y \in [0,1].$$

Under Lemma 5, it holds that $\underline{b}_\mu \leq \underline{b}_\nu$ and $\overline{b}_\mu \geq \overline{b}_\nu$. This allows us to check the inequality separately on the three intervals $[0, \underline{b}_\nu]$, $\left[\underline{b}_\nu, \overline{b}_\nu\right]$ and $\left[\overline{b}_\nu, 1\right]$.

For all $y \in [0, \underline{b}_\nu]$, it holds that

$$\int_0^y G^\gamma(x)\mathrm{d}x = 0 \leq \int_0^y F^\gamma(x)\mathrm{d}x.$$

For all $y \in \left[\underline{b}_\nu, \overline{b}_\nu\right]$ it holds that

$$\begin{aligned}
\int_0^y G^\gamma(x)\mathrm{d}x &= \int_{\underline{b}_\nu}^y \gamma(0) + \gamma(\lambda_0)G(x)\mathrm{d}x \\
&= \gamma(0)(y - \underline{b}_\nu) + \gamma(\lambda_0)\int_0^y G(x)\mathrm{d}x - \gamma(\lambda_0)\int_0^{\underline{b}_\nu} G(x)\mathrm{d}x
\end{aligned}$$

$$
\begin{aligned}
&= \gamma(0)(y-p_0) + \gamma(\lambda_0)\int_0^y G(x)\mathrm{d}x \\
&\leq \gamma(0)(y-p_0) + \gamma(\lambda_0)\int_0^y F(x)\mathrm{d}x \\
&= \int_0^y F^\gamma(x)\mathrm{d}x.
\end{aligned}
$$

The first equality follows from the definition of virtual posteriors and the equilibrium strategies. For the third equality, we apply Lemma 4. The inequality follows from the assumption that $\mu \succ \nu$.

Since $G^\gamma(x) = 1$ for $x \geq \bar{b}_\nu$, it follows that for all $y \in \left[\bar{b}_\nu, 1\right]$ it holds that

$$
\int_y^1 G^\gamma(x)\mathrm{d}x \geq \int_y^1 F^\gamma(x)\mathrm{d}x.
$$

The expected value of both distributions is consistent with the common prior, i.e.,$\int_0^1 F^\gamma(x)\mathrm{d}x = \int_0^1 G^\gamma(x)\mathrm{d}x = 1 - p_0$. Thus, we conclude that

$$
\int_0^y G^\gamma(x)\mathrm{d}x \leq \int_0^y F^\gamma(x)\mathrm{d}x,
$$

for all $y \in [\bar{b}_\nu, 1]$, which concludes the proof. $\square$

Proof of Theorem 2. Under known partisanship, it is common knowledge between partisan senders and the receiver that the partisan senders will choose whatever message gets higher weight for their preferred action. Hence, the receiver knows that no message send by a partisan is informative, and hence the virtual posterior derived from a partisan sender is always the prior. In contrast, the virtual posterior distribution of known advisors in the most informative equilibrium is their posterior distribution, as in Proposition 1. We hence arrive at the following virtual posterior distribution for known partisanship:

$$
\mu^\gamma_{known} = \gamma(\lambda_0)\mu(x) + (\gamma(0) + \gamma(1))\,\delta_{p_0}
$$

with average intensity:

$$
\pi(\mu_{known}) = \gamma(\lambda_0)\pi(\mu)
$$

The virtual posterior distribution derived from the equilibrium in Theorem 1 is given by

$$
\mu^{\gamma}_{unknown}(x) = \begin{cases} \gamma(\lambda_0)F(x) + \gamma(0) & \text{if } x = \underline{b} \\ \gamma(\lambda_0)\mu(x) & \text{if } x \in \left(\underline{b}, \overline{b}\right) \\ \gamma(\lambda_0)\left(1 - F(x)\right) + \gamma(1) & \text{if } x = \overline{b} \\ 0 & \text{else} \end{cases}
$$

Let us first check that $\mu_{unknown}$ is less informative then μ_{known}. For this, let us denote with $F^{\gamma}_{unknown}$, F^{γ}_{known}, and F the cdfs of $\mu^{\gamma}_{unknown}$, μ^{γ}_{known} and μ respectively. We then have to show that

$$
\int_0^y F_{known}(x)\mathrm{d}x \geq \int_0^y F_{unknown}(x)\mathrm{d}x \ \forall y \in [0,1].
$$

When $y \in [0, \underline{b}]$, this is true, since $F_{unknown}$ is constant and equal to 0 on this interval. The case for the interval $y \in [\overline{b}, 1]$ follows by a symmetric argument, since the integrals become equal to the prior at $y = 1$. Let us hence focus on $y \in [\underline{b}, \overline{b}]$. We then get

$$
\begin{aligned}
\int_0^y F_{known}(x)\mathrm{d}x &= \int_0^y \gamma(\lambda_0)F(x) + (1-\gamma(\lambda_0))\mathbb{1}\{x \geq p_0\})\mathrm{d}x \\
&= \int_0^{\underline{b}} \gamma(\lambda_0)F(x)\mathrm{d}x + \int_{\underline{b}}^y \gamma(\lambda_0)F(x) + (1-\gamma(\lambda_0))\mathbb{1}\{x \geq p_0\})\mathrm{d}x \\
&\geq \gamma(0)(p_0 - \underline{b}) + \int_{\underline{b}}^y \gamma(\lambda_0)F(x) + \gamma(0)\mathbb{1}\{x \geq p_0\})\mathrm{d}x \\
&\geq \int_{\underline{b}}^y \gamma(\lambda_0)F(x) + \gamma(0)\mathrm{d}x \\
&= \int_0^y F_{unknown}(x)\mathrm{d}x.
\end{aligned}
$$

The first inequality makes use of Lemma 4. In the last equality, we use the fact that $F_{unknown}$ is equal to 0 on $[0, \underline{b}]$. Lastly, we verify the average informativeness of $\mu_{unknown}$ to be

$$
\pi(\mu^{\gamma}_{unknown}) = \int_0^1 |x - p_0|\mathrm{d}\mu_{unknown}
$$

$$=\gamma(\lambda_0)\int_0^1 |x-p_0|\mathrm{d}\mu + \gamma(0)(p_0-\underline{b}) + \int_0^{\underline{b}} (\underline{b}-x)\mathrm{d}\mu + \gamma(1)(\overline{b}-p_0) + \int_{\overline{b}}^1 (x-\overline{b})\mathrm{d}\mu$$
$$=\gamma(\lambda_0)\pi(\mu).$$

Here the last equation is a consequence of Lemma 4.

□

Proof of Proposition 3. We prove the proposition in two steps. We start to show that by monotonicity and continuity of $\underline{b}$ and $\overline{b}$, there exists $c_0, c_1 \in (0,1)$, such that the receiver can only form two expected posteriors in the optimal equilibrium. Then, we prove that there exist c_0 and c_1 such that $c_0 + c_1 < 1$. For both parts, we use Lemma 4, which characterizes the lower expertise bound by the equation

$$\gamma(0)\left(p_0-\underline{b}\right) = \gamma(\lambda_0)\cdot\int_0^{\underline{b}} F(x)\mathrm{d}x.$$

The lower expertise bound can take any value in $\underline{b} \in [0, \max\{x : F(x) = 0\}]$ if $\gamma(0) = 0$. Further, it is p_0 if $\gamma(0) = 1$. Rewriting the above equation yields

$$\frac{\gamma(0)}{\gamma(\lambda_0)} = \frac{\int_0^{\underline{b}} F(x)\mathrm{d}x}{p_0-\underline{b}} \tag{2}$$

which exhibits that $\underline{b}$ is monotonically increasing in $\gamma(0)$, monotonically decreasing in $\gamma(\lambda_0)$ and continuous in $\gamma(0), \gamma(\lambda_0) \in (0,1)$.

Since μ is never-ignorant, there exists a highest type strictly smaller than the prior, $p_L := \max\{x | x < p_0 \wedge x \in \mathcal{P}\}$. The proposition is fulfilled if the lower expertise bound equals this type $\underline{b} = p_L$. Continuity and monotonicity of $\underline{b}$ imply that the right-hand side of Equation (2) is positive and finite, and hence $\gamma(0) < 1$ if $\underline{b} = p_L$. The proof for the upper part with type $p_H := \min\{x | x > p_0 \wedge x \in \mathcal{P}\}$ is analogous, so that constants $c_0, c_1 \in (0,1)$ are implicitly given by

$$c_0\left(p_0-p_L\right) = \gamma(\lambda_0)\cdot\int_0^{p_L} F(x)\mathrm{d}x \quad \text{and} \quad c_1\left(p_H-p_0\right) = \gamma(\lambda_0)\cdot\int_{p_H}^1 1-F(x)\mathrm{d}x. \tag{3}$$

To see that $c_0 + c_1 < 1$ divide Equations (3) by $(p_0 - p_L)$ and $(p_H - p_0)$, respectively. Adding both equations yields

$$c_0 + c_1 = \gamma(\lambda_0)\cdot\frac{\int_0^{p_L} F(x)\mathrm{d}x}{p_0-p_L} + \gamma(\lambda_0)\cdot\frac{\int_{p_H}^1 1-F(x)\mathrm{d}x}{p_H-p_0}.$$

Since $\frac{\int_0^{p_L} F(x)\mathrm{d}x}{p_0-p_L}, \frac{\int_{p_H}^1 1-F(x)\mathrm{d}x}{p_H-p_0} > 0$, it follows that $\gamma(\lambda_0) > 0$. This implies that $c_0 + c_1 = \gamma(0) + \gamma(1) = 1 - \gamma(\lambda_0) < 1$, which completes the proof. □

Proof of Proposition 4. Let $c_0^\mu, c_0^\nu \in (0,1)$ constants from Proposition 3 for distributions μ and ν with cdfs F and G, respectively. Define $c_0 = \max\{c_0^\mu, c_0^\nu\}$ as the smallest constant, such that both virtual posteriors μ^γ and ν^γ have only two mass points. Now we compare the resulting lower expertise bounds $\underline{b}^{\mu^\gamma}$ and $\underline{b}^{\nu^\gamma}$. The smaller the lower expertise bound, the better the signal. Thus, it is sufficient to show that $\underline{b}^{\mu^\gamma} < \underline{b}^{\nu^\gamma}$.

Under Proposition 3 and Lemma 4, it follows that

$$\underline{b}^{\mu^\gamma} = p_0 - \frac{\gamma(\lambda_0)}{\gamma(0)} \int_0^{p_0} F(x)\mathrm{d}x,$$

Lemma 6 implies that $\underline{b}^{\mu^\gamma} < \underline{b}^{\nu^\gamma}$ which completes the proof for the lower expertise bound.

The proof for the upper expertise bounds is analogous. Further, $c_0 + c_1 < 1$ remains to be true despite taking the maximum and minimum, respectively, by the same reason as in the proof of Proposition 3. □

Lemma 6. *Let μ and ν be posterior distributions with cdfs F and G, respectively. Then, $\pi(\mu) > \pi(\nu)$ if and only if*

$$\int_0^{p_0} F(x)\mathrm{d}x > \int_0^{p_0} G(x)\mathrm{d}x \quad \text{and} \quad \int_{p_0}^1 1 - F(x)\mathrm{d}x > \int_{p_0}^1 1 - G(x)\mathrm{d}x.$$

Proof. We rewrite $\pi(\mu)$ until we arrive at an expression from which the result is immediate:

$$\begin{aligned}
\pi(\mu) &= \mathbb{E}\left[|p_i - p_0|\right] \\
&= \int_0^1 |x - p_0| \; \mu(x)\mathrm{d}x \\
&= \int_0^{p_0} (p_0 - x) \; \mu(x)\mathrm{d}x + \int_{p_0}^1 (x - p_0) \; \mu(x)\mathrm{d}x \\
&= p_0 \cdot F(p_0) - \int_0^{p_0} x \; \mu(x)\mathrm{d}x + \left(p_0 - \int_0^{p_0} x \; \mu(x)\mathrm{d}x\right) - p_0 \cdot (1 - F(p_0)) \\
&= 2\left(p_0 \cdot F(p_0) - \int_0^{p_0} x \; \mu(x)\mathrm{d}x\right) \\
&= 2\int_0^{p_0} F(x)\mathrm{d}x.
\end{aligned}$$

The fourth equation follows from the common prior $p_0 = \int_0^1 x \; \mu(x)\mathrm{d}x$ and the last equality from integration by parts. The derivation of the second inequality is analogous. □

B Appendix: Comparison of Knowledge Distributions with Blackwell, Rothschild and Stiglitz

In this part of the Appendix, we collect certain tools from the literature that we use throughout and that are related to definition 2. This allows us to compare the receiver's utility for different posterior distributions of senders. The methods and results in this subsection are borrowed from Chapter 12 in Blackwell and Girshick [1979] and Rothschild and Stiglitz [1970]. In order to apply their machinery to our problem, we adjust our setting, and translate our notation into theirs.

The following results rely on the assumption that the action space of the receiver is a closed bounded convex subset of $\mathbb{R}$. To fulfill this assumption, we extend the action space of the receiver from $\{0, 1\}$ to $\Delta\{0, 1\}$, so that her action space is the interval $[0, 1]$. An action $a \in \Delta\{0, 1\}$ corresponds to the probability that the receiver takes action 1. Note that we can use this extended action space throughout the whole paper without changing any result. In all statements on best responses of the receiver, one of the two extreme actions $\{0, 1\} \subset \Delta\{0, 1\}$ is optimal. We use the action space $\{0, 1\}$ in the main text of the paper to simplify the exposition.

To present the next results, it is also helpful to introduce some of the notation of Blackwell and Girshick [1979][8] For a posterior distribution μ, we define a $2 \times N$ matrix P, where $N = |\mathcal{P}|$ is the number of possible posteriors. The rows represent the two states of the world 0 and 1. Each column represents one possible posterior. The value P_{ij} is the probability of observing the posterior represented by column j in state i. Note that matrix P is Markov, which means that $P_{ij} > 0$ for all i and j and that $\sum_{j=1}^{N} P_{ij} = 1$ for all i. With the notation, we are equipped to remind the reader of Theorem 12.2.2 in Blackwell and Girshick [1979].

Proposition A (Blackwell and Girshick [1979]). *Let P and Q be two $2 \times N_1$ and $2 \times N_2$ Markov matrices of posterior distributions μ and ν. μ is more informative than ν if and only if there is an $N_1 \times N_2$ Markov matrix M with $PM = Q$.*

Matrix M is said to *garble* information by transforming matrix P to Q. This means that distribution ν can be constructed from distribution μ. This interpretation justifies the statement that μ is more informative than ν.

The next result generalizes the previous proposition by enabling the comparison of sets of distributions. Each sender sends a conditionally independent posterior. Consider two sets of senders with different posterior distributions. Then, Theorem 12.3.2 in Blackwell

[8]We also enjoyed reading the notes of Borgers [2009] on Chapter 12 of Blackwell and Girshick [1979] and borrow some of his notation.

and Girshick [1979] allows us to compare the information of both groups in the following sense.

Proposition B (Blackwell and Girshick [1979])**.** *Let* $(\mu_i)_{i=1}^n$ *and* $(\nu_i)_{i=1}^n$ *be two sets of posterior distributions. Suppose that* μ_i *is more informative than* ν_i *for every* i*. Then, the combination of posterior distributions* $(\mu_i)_{i=1}^n$ *is more informative than* $(\nu_i)_{i=1}^n$*.*

The proposition allows, to compare the information that is transmitted to the receiver from different distributions. Theorem 12.2.2 (4) in Blackwell and Girshick [1979] allows us to use this result for a statement on the utility of the receiver.

Proposition C (Blackwell and Girshick [1979])**.** *Let* μ *and* ν *be two posterior distributions such that* μ *is more informative than* ν*. Then, for every continuous convex function* $\phi : [0,1] \to \mathbb{R}$ *we have*

$$\mathbb{E}_{\mu}\left[\phi(x)\right] \geq \mathbb{E}_{\mu'}\left[\phi(x)\right].$$

Note that the utility function $u^{\star}(q)$ is convex in q. Thus, if there are two posterior distributions with $\mu \succ \nu$, the proposition implies that the expected utility for the receiver with distribution μ is at least as high as with distribution ν.

Lastly, we state a result by Rothschild and Stiglitz [1970] that links second-order stochastic dominance to mean-preserving spreads in a way more suitable for our needs, than a similar result by Blackwell and Girshick [1979].

Proposition D (Rothschild and Stiglitz [1970])**.** *Let* p *and* p' *be random posteriors drawn from the measures* μ *and* ν*, respectively. Then* $\mu \succ \nu$ *if and only if there exists a random variable* z *s.t.*

$$p \overset{d}{=} p' + z$$

and

$$E[z|p'] = 0 \ \forall p'.$$

References

Ricardo Alonso, Wouter Dessein, and Niko Matouschek. When does coordination require centralization? *American Economic Review*, 98(1):145–79, 2008.

David Austen-Smith. Information transmission in debate. *American Journal of Political Science*, pages 124–152, 1990.

Yaron Azrieli. The price of 'one person, one vote'. *Social Choice and Welfare*, 50(2): 353–385, 2018.

Michael R Baye, Dan Kovenock, and Casper G De Vries. Rigging the lobbying process: an application of the all-pay auction. *The American Economic Review*, 83(1):289–294, 1993.

David Blackwell. Comparison of experiments. Technical report, Howard University, Washington United States, 1950.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, pages 265–272, 1953.

David A Blackwell and Meyer A Girshick. *Theory of Games and Statistical Decisions*. Courier Corporation, 1979.

Tilman Borgers. Notes on blackwell's comparison of experiments, 2009. URL `http://www-personal.umich.edu/~tborgers/Blackwell-Girshick.pdf`.

James M Buchanan, Robert D Tollison, and Gordon Tullock. *Toward a theory of the rent-seeking society*. Number 4. Texas A & M Univ Pr, 1980.

Marquis de Condorcet. Essay on the application of analysis to the probability of majority decisions. *Imprimerie Royale Paris*, 1785.

Vincent P Crawford and Joel Sobel. Strategic information transmission. *Econometrica*, pages 1431–1451, 1982.

Timothy Feddersen and Wolfgang Pesendorfer. Voting behavior and information aggregation in elections with private information. *Econometrica*, pages 1029–1058, 1997.

Timothy J Feddersen and Wolfgang Pesendorfer. The swing voter's curse. *The American Economic Review*, pages 408–424, 1996.

Dino Gerardi, Richard McLean, and Andrew Postlewaite. Aggregation of expert opinions. *Games and Economic Behavior*, 65(2):339–371, 2009.

Thomas W Gilligan and Keith Krehbiel. Asymmetric information and legislative rules with a heterogeneous committee. *American Journal of Political Science*, pages 459–490, 1989.

Jacob Glazer, Helios Herrera, and Motty Perry. The market for product reviews. Working paper, 2019.

Patrick Hummel, John Morgan, and Phillip C Stocken. A model of flops. *The RAND Journal of Economics*, 44(4):585–609, 2013.

Vijay Krishna and John Morgan. A model of expertise. *The Quarterly Journal of Economics*, 116(2):747–775, 2001.

Ming Li and Kristóf Madarász. When mandatory disclosure hurts: Expert advice and conflicting interests. *Journal of Economic Theory*, 139(1):47–74, 2008.

Andrew McLennan. Consequences of the condorcet jury theorem for beneficial information aggregation by rational agents. *American Political Science Review*, 92(2):413–418, 1998.

Joseph McMurray. Ideology as opinion: a spatial model of common-value elections. *American Economic Journal: Microeconomics*, 9(4):108–40, 2017.

Shmuel Nitzan and Jacob Paroush. Optimal decision rules in uncertain dichotomous choice situations. *International Economic Review*, pages 289–297, 1982.

Charles Piller. Hidden conflicts? *Science*, 361(6397):16–20, 2018. ISSN 0036-8075. doi: 10.1126/science.361.6397.16. URL `http://science.sciencemag.org/content/361/6397/16`.

Michael Rothschild and Joseph E Stiglitz. Increasing risk: I. a definition. *Journal of Economic Theory*, 2(3):225–243, 1970.

Asher Wolinsky. Eliciting information from multiple experts. *Games and Economic Behavior*, 41(1):141–160, 2002.