The stability of expert panels in persuasion games

Igor Karpov*

October 14, 2026

Abstract

Can adding an expert to a panel make a decision-maker worse off, even though the expert can only reveal information? I study a persuasion game in which a decision-maker consults a panel of experts who have state-independent agendas, may be uninformed, and can only disclose or conceal verifiable evidence. A new expert reveals some states that were concealed, but he also shifts the decision-maker's default action, the action she takes when no expert reports, and the other experts adjust what they reveal. I show that a panel of sufficiently extreme experts is stable: no additional expert, whatever his agenda and quality, makes the decision-maker worse off. A panel of moderate experts is generically unstable: for every quality level, some expert close to the default action shifts it so that an incumbent stops revealing states far from it. With commitment to the default action every panel is stable, and extreme panels behave as if the decision-maker could commit. Polarized committees can therefore be expanded safely, while adding a centrist to a moderate committee can backfire.

Keywords: Persuasion games, Verifiable disclosure, Expert panels, Committee design

JEL Classification: C72, D71, D82, D83

*University of Chicago. This paper is based on my undergraduate thesis (Joint Bachelor's Programme in Economics of HSE and NES, Moscow, 2020). I am very grateful to my supervisor, A. V. Savochkin, for his guidance and support. All errors are my own.

1 Introduction

When making decisions, policymakers, CEOs and ministers often rely on the opinions of the experts they consult, and the literature (e.g. Milgrom, 1981; Grossman, 1981; Kamenica and Gentzkow, 2011) has recognized that these experts have their own preferences and incentives to reveal information strategically. Suppose that the government is contemplating a policy on poverty and has formed a panel of experts to advise it on how strict the regulation should be. An expert may be economically left-wing and prefer strong support for the poor, or right-wing and prefer to remove any regulation, and these inclinations are often publicly known from past public service or institutional affiliations: an expert may be a lobbyist for a parliamentary party that stands to lose if the regulation is passed, or a social activist who has publicly argued for it. Common sense suggests that the government should add another expert whenever one is available: an additional expert can only add information, and experts with different agendas preclude each other's attempts to hide unfavorable evidence. Can an additional expert, who can only reveal what he knows, make the decision-maker worse off?

In this paper, I study this question in the persuasion game of Bhattacharya and Mukherjee (2013, hereafter BM) and Bhattacharya et al. (2018, hereafter BGM). A decision-maker wants her action to match an unknown state. Each expert is informed with some probability, his quality, and has a state-independent ideal action, his agenda. Information is verifiable: an uninformed expert must admit ignorance, while an informed expert can either reveal the state or pretend to be uninformed. The decision-maker cannot commit, so when no expert reports she takes a default action that is optimal given what the experts reveal, and each expert reveals exactly the states that are closer to his agenda than the default action. BM and BGM ask which panel of a given size is best. I take the panel as given and ask whether adding one expert with a given agenda and quality benefits the decision-maker. This marginal question is the relevant one when she cannot choose freely from a continuum of experts but must decide whether to add the expert who is available.

A new expert affects the decision-maker in two ways. The *direct effect* is that he reveals states that were previously concealed, which always benefits her. The *strategic effect* is that his disclosure shifts the default action, and the experts already in the panel respond by changing what they reveal. Its sign depends on where the revelation sets end. Each expert's revelation set is an interval with one end at the default action and the other end at the state that lies equally far from his agenda on the opposite side. For an extreme expert this far end is the boundary of the state space and does not move, so a shift of the default action only passes the states around it from one expert to another. For a moderate expert the far end moves as well, and a shift of the default action can make him stop revealing states far from it.

Extreme panels. The first main result, Theorem 1, is that a panel of sufficiently extreme experts is stable: an additional expert, whatever his agenda and quality, never makes the decision-maker worse off, in any equilibrium of the enlarged panel. For an extreme panel the same holds when several experts are added at once, and both statements hold for any prior and any loss function in the class considered here. The reason is commitment. Since the only end of an extreme expert's revelation set that moves is the default action itself, where the decision-maker's loss is zero, the equilibrium default action is exactly the one she would choose if she could commit to it, and at that action a new expert can only add disclosure.

Moderate panels. The second main result, Theorem 2, is that a panel of two moderate experts, whose revelation sets do not reach the boundaries of the state space, is in general unstable. What matters is the decision-maker's marginal value of commitment: the change in her payoff from a small shift of the default action when the experts' revelation sets respond to it. Whenever this value is not zero, for every quality level there is an expert whose addition harms her: an expert whose agenda lies close to the default action, on the side from which he pushes it in the direction opposite to the one she would commit to. His own disclosure is small, but one of the incumbents stops revealing some states far from the default

action, and the information lost outweighs the information gained. The condition cannot be dropped: when the marginal value of commitment is zero, as in every symmetric panel with a symmetric prior, the strategic effect vanishes to first order, and, under regularity conditions, the panel is stable with respect to every expert of sufficiently low quality (Proposition 1).

Which experts are harmful. Proposition 2 characterizes the harmful experts. The change in the decision-maker's utility caused by a new expert equals the change in the payoff she would obtain by committing to the new rather than the old default action, plus the direct effect of the new expert, and the direct effect is bounded by the distance between his agenda and the default action he induces, times how strongly he must push it. Hence a harmful expert must move the default action to one that would give her a lower payoff under commitment, and experts that are narrow but well informed are harmful: they disclose little, but they can move the default action a long way. In the leading example of the paper, a harmful expert pushes the default action towards the better informed incumbent, who then stops revealing some extreme states on his side (Example 5).

New equilibria. A new expert matters also through the set of equilibria. He can create new equilibria even when the initial panel has a unique one, and the new equilibria can be worse than the initial one even when the equilibrium that continues it improves; this can happen even for an expert with an extreme agenda (Examples 3 and 4).

Commitment. When the decision-maker can commit to her default action, every panel is stable (Proposition 3). Together with the two theorems, this gives the organizing observation of the paper: extreme panels behave as if the decision-maker could commit, and a new expert can harm a panel only by moving the default action to one at which her payoff under commitment is lower. The practical reading is that a committee of radical experts can be expanded without risk, while adding a centrist voice to a committee of moderates can backfire, because it changes what the decision-maker does when no one speaks, and thereby

what the incumbents are willing to reveal.

1.1 Related literature

As mentioned above, the framework I study closely follows that of BM and BGM. The first main result of BM is that the best way to compose a panel of experts with fixed qualities is to select extreme agendas. The second main result of BM is that with sufficiently extreme experts an increase in the quality of any one of them strictly increases the utility of the decision-maker. Hence, restricting attention to extreme panels with agendas in $\{0, 1\}$ makes the addition of every new expert with an extreme agenda beneficial for the decision-maker, since it is equivalent to an increase in the quality of one of the existing experts. BM also show that outside this class an increase in the quality of an expert can reduce the decision-maker's payoff; since adding an expert with the agenda of an existing one is equivalent to increasing that expert's quality, this already implies that an additional expert can be harmful. In the subsequent paper, BGM investigate the extreme subframework in depth in order to determine the conditions under which a homogeneous (in agendas) or a heterogeneous panel is optimal. Gong and Yang (2018) study the same comparison in a discrete model of disclosure in which a single piece of evidence can arise in several states and experts can disclose partially. They identify a burden-of-proof effect, which favors homogeneous panels, and a balance-of-opinions effect, which favors diverse ones, and show that diverse panels are optimal if and only if experts are relatively well-informed. Both papers compare panels of a given size but different composition, whereas this paper studies the complementary question of adding a new expert to a fixed panel rather than composing a new one.

The model of BM and BGM belongs to the broad literature on strategic communication games. Starting with Milgrom (1981), a vast literature on persuasion games has developed over the last few decades. Some of the primary contributions to this literature, such as Milgrom (1981), Grossman (1981) and Milgrom and Roberts (1986), argue that with verifiable messages a skeptical decision-maker can extract all relevant information in equilibrium

if the expert is known to possess it. Several authors have also noted that this so-called full-revelation result is rather sensitive to assumptions about the expert's information: if there is a positive probability that the expert does not possess any relevant information, then an expert who obtains adverse information can hide it in equilibrium through selective disclosure (see, e.g., Dye, 1985; Okuno-Fujiwara et al., 1990; Shavell, 1994; Shin, 1994; Jung and Kwon, 1988).

The cheap talk literature has also studied information revelation with multiple senders. Gilligan and Krehbiel (1989) and Austen-Smith (1990, 1993) provide early cheap talk models that explore the informational properties of "debates" between multiple agents who possess some information but have conflicting interests. As an example of a more recent contribution, Krishna and Morgan (2001) explore the value of competition (i.e., experts with divergent biases) in dynamic communication in a unidimensional state/action space. In related work, Battaglini (2002) shows that if the state/action space is instead multidimensional and unbounded, then there exists an equilibrium in which the true state is always revealed. Whereas the cheap talk literature usually considers settings in which the experts cannot be uninformed about the state, both BM and the present paper, like the persuasion papers mentioned above, are interested in situations with uncertainty about the experts' knowledge. In this model, when the experts have completely opposed agendas, information is fully revealed to the decision-maker only in the event that both experts know the state. This result is close to the full-revelation result with opposed biases in Krishna and Morgan (2001). However, BM note that this uncertainty about the experts' information creates new channels of strategic manipulation that the decision-maker has to contend with. Moreover, competition between experts may limit the ability of the decision-maker to induce all experts to reveal information simultaneously.

While a significant number of recent papers in this field focus on optimal rules of persuasion (Glazer and Rubinstein, 2004, 2006; Kamenica and Gentzkow, 2011), the optimal composition of expert panels in persuasion games has not been studied as extensively. Still,

a few papers compare the efficiency of a panel of multiple experts with the case in which a single expert gathers all the information. For example, Shin (1998) compares a panel of two experts with opposite biases with one unbiased expert and demonstrates that the former may indeed reveal more information. Dewatripont and Tirole (1999) make a similar comparison, but their setting differs in that it includes costly information acquisition and transferable utility. These results are somewhat predictable: the use of opposing experts (or "advocacy") creates sharper incentives for information acquisition, hence two experts are better than one. More recently, Gentzkow and Kamenica (2017) present a general setting in which the experts commit to an information structure ex ante, which eliminates any information asymmetry between the experts and the decision-maker. They show that greater divergence between the experts' preferences results in more information being revealed in equilibrium, when the comparison is made appropriately. Notably, all of the aforementioned papers conclude that a heterogeneous panel is optimal for the decision-maker.

However, the literature gives no single answer to this question. For instance, Gerardi and Yariv (2008) consider a model with binary state and action and show that the optimal panel consists of experts with identical extreme preferences, which directly contradicts the results above. Again, their model is quite different from BM, in the sense that the experts acquire information at a cost, and after the acquisition the information becomes publicly known. Therefore, the construction of the optimal panel is mostly a problem of providing optimal incentives for information acquisition. An intermediate position is taken by Beniers and Swank (2004), who consider a rather specific setting in which the experts can acquire both verifiable and unverifiable information and, unlike in BM, have state-dependent preferences. They show that the optimal committee should consist of similar experts whose preferences are close to those of the decision-maker if the cost of information acquisition is low, and vice versa. The results of BGM are not unambiguous either: they investigate conditions under which a panel of extreme opposing experts is preferred to a panel of extreme identical ones and vice versa, and do not argue in favor of either, because the conditions for both panels

are realistic enough and neither panel has any special features.

The strand of literature closest to this paper compares the effectiveness of committees of different sizes. In the BGM model, consulting an additional expert always strictly improves the welfare of the decision-maker; however, this conclusion need not hold in general. Krishna and Morgan (2001) and Kartik et al. (2017) demonstrate that in settings with unverifiable messages or with costly information acquisition, respectively, the addition of a new expert can harm the decision-maker. Cai (2009) considers a model of costly information acquisition with heterogeneous preferences, in which the optimal number of experts depends on the structure of the panel. The optimal number is pinned down within the model, and the author provides an equation for it that stems from the optimization problem. These results are similar to mine in some sense, but none of these settings combines the two features of BM that are important for its main result: verifiability of the messages and costless information acquisition. Using the BM setting, this paper focuses on a situation in which each expert can only reveal the truth or conceal it, without partial revelation. So the strategic concealment of verifiable information is the only channel of information manipulation, and I abstract from the strategic issues associated with information acquisition and its costs.

2 The Model

2.1 Setup

The model closely follows the general environment of BM, with some refinements from BGM.[1] Consider a decision-maker who must take an action $y \in Y = [0, 1]$ based on information about the underlying state of the world $\theta \in \Theta = [0, 1]$. The decision-maker wants to choose an action y that minimizes her loss, which depends only on the distance between her action and the optimal action $y_J(\theta)$. As a normalization, I assume without loss of generality that the latter equals the state of the world, $y_J(\theta) = \theta$, and hence the decision-maker's payoff is

[1] For example, the support of the state and of the action is the unit interval.

$u_J(y,\theta) = -v(|y-\theta|)$. I assume that $v(0) = v'(0) = 0$ and $v'(x) > 0$, $v''(x) > 0$ for all $x > 0$. The decision-maker does not observe the realized state; she only knows that it is distributed according to a prior with CDF $F(\theta)$. I assume that F has a continuous density $f(\theta)$, which is positive on $[0,1]$.

The decision-maker can solicit information from a panel of n experts A_i, $i \in \{1,\dots,n\}$. Each expert A_i has a private type $\tau_i \in \{0,1\}$. If $\tau_i = 1$, the expert observes the state; if $\tau_i = 0$, he observes nothing. I assume that the experts' types are independent, and let $a_i = \Pr(\tau_i = 1) \in (0,1)$ be the quality of expert A_i. The payoff of expert A_i from action y is given by $u_i(y,x_i) = g_i(|y-x_i|)$ for some strictly decreasing function g_i and an individual state-independent ideal action $x_i \in [0,1]$, called the agenda of expert A_i. Thus the payoff of expert A_i is symmetric and single-peaked around x_i and strictly decreasing in the distance between his agenda and the decision-maker's action.

Following BM, I refer to a tuple of an agenda profile $(x_1,\dots,x_n) \in [0,1]^n$ and a corresponding quality profile $(a_1,\dots,a_n) \in (0,1)^n$ as a panel. Before the decision-maker chooses her action, each expert sends her a message m_i. The messages are verifiable. An uninformed expert must admit that he is uninformed, i.e., report $m_i = \emptyset$, but an informed expert has the option of either revealing the true state or feigning ignorance and pooling with the uninformed type; i.e., the message set of an informed expert in state θ is $M = \{\theta, \emptyset\}$. Given a profile of messages $m = (m_1,\dots,m_n)$, the decision-maker takes an action $y(m) \in [0,1]$. Like BM, I use perfect Bayesian equilibrium as the solution concept, and, since F is nonatomic, I restrict attention to pure strategies without loss of generality.

2.2 Equilibrium

In what follows, I briefly state the key properties of the equilibrium and elaborate mainly on those that play a key role in the analysis of the optimal panel.

In equilibrium, if at least one expert reveals the state (i.e., if $m_i = \theta$ for some $i \in \{1,\dots,n\}$), then the decision-maker chooses $y(m) = \theta$. Thus the decision-maker's strategy

is completely characterized by the action $y^* = y(\emptyset, \ldots, \emptyset)$ that she takes when no expert reports the state. Strict concavity of u_J in y makes randomization suboptimal for the decision-maker, hence her optimal strategy is deterministic. Following BM, I call y^* the decision-maker's *default action*. The strategy of each informed expert A_i is described by his revelation set $\Theta_i = \{\theta \in \Theta \mid m_i = \theta\}$, the set of states that he reports truthfully. An uninformed expert must admit ignorance, i.e., $m_i = \emptyset$ if $\tau_i = 0$.

Note that an informed expert's choice of message matters only when no other expert reports, i.e., if $m_{-i} = (\emptyset, \ldots, \emptyset)$. Conditional on this event, expert A_i induces the action $y = \theta$ by disclosing the state and obtains the default action $y = y^*$ if he chooses not to disclose. Thus, given a default action y^*, expert A_i reports the state if and only if θ is closer to his ideal point x_i than y^* is. In other words, revealing the state is a best response for informed expert A_i if and only if

$$\theta \in \Theta_i(y^*) = \{\theta \in \Theta : |\theta - x_i| \leq |y^* - x_i|\}, \tag{1}$$

i.e., $\Theta_i(y^*) = [2x_i - y^*, y^*] \cap [0,1]$ if $y^* \geq x_i$ and $\Theta_i(y^*) = [y^*, 2x_i - y^*] \cap [0,1]$ if $y^* \leq x_i$.

Given the revelation sets, state θ remains undisclosed with probability

$$P(\theta; y^*) = \prod_{i=1}^{n} \left(1 - a_i \mathbf{1}\{\theta \in \Theta_i(y^*)\}\right).$$

The default action must be optimal for the decision-maker given this probability, and since u_J is strictly concave in y, y^* is an equilibrium default action if and only if

$$V(y^*) \equiv \int_0^1 P(\theta; y^*)\, u_J'(y^*, \theta)\, dF(\theta) = 0, \tag{2}$$

where here and below u_J' and u_J'' denote derivatives with respect to the action y. The

expected utility of the decision-maker in this equilibrium is

$$\mathbb{E}[U] = \int_0^1 P(\theta; y^*)\, u_J(y^*, \theta)\, dF(\theta). \tag{3}$$

Since V is continuous, with $V(0) > 0 > V(1)$, an equilibrium always exists, but it need not be unique.

Example 1. Let $n = 2$, $x_1 = x_2 = 0$ and $a_1, a_2 \in (0, 1)$. Then there exists a unique equilibrium with $\Theta_1^* = \Theta_2^* = [0, y^*]$, where y^* solves

$$(1 - a_1)(1 - a_2) \int_0^{y^*} u_J'(y^*, \theta) dF(\theta) + \int_{y^*}^1 u_J'(y^*, \theta) dF(\theta) = 0.$$

This equilibrium exists and is unique: the first term of the sum is negative and the second is positive, both are decreasing in y^* (the boundary terms vanish because $u_J'(y^*, y^*) = 0$), and the left-hand side is positive at $y^* = 0$ and negative at $y^* = 1$; hence, by continuity of u_J', the equation has a unique solution in $(0, 1)$.

3 Stability of expert panels

3.1 Examples and definition

I begin by showing, using simple examples, that the direction of the change in the decision-maker's utility after the addition of a new expert is ambiguous.

Example 2. Let the prior be uniform and $u_J(y, \theta) = -(y - \theta)^2$, and consider a panel of two experts with $x_1 = 0.4$, $a_1 = 0.1$, $x_2 = 0.6$, $a_2 = 0.9$. Its equilibrium has $y^* \approx 0.471$ and $\mathbb{E}[U] \approx -0.07893$. If we add a third expert with agenda $x_3 = 0.45$ and quality $a_3 = 0.5$, the equilibrium becomes $y^* \approx 0.472$ and $\mathbb{E}[U] \approx -0.07897$, so the utility of the decision-maker decreases.

In Sections 4 and 5 I describe the main reasons why the utility decreases in this example

and increases in the examples below. Both panels in Example 2 have a unique equilibrium, so the example does not depend on equilibrium selection. A new expert can, however, also change the number of equilibria.

Example 3. In the panel of Example 2, add instead an expert with agenda $x_3 = 0.55$ and quality $a_3 = 0.5$. The enlarged panel has three equilibria. In the one that continues the initial equilibrium, $y^* \approx 0.4698$ and $\mathbb{E}[U] \approx -0.07879$: the new expert moves the default action slightly to the left and makes the decision-maker better off. In the other two, $y^* \approx 0.1968$ and $y^* \approx 0.2013$, with $\mathbb{E}[U] \approx -0.0138$ and $\mathbb{E}[U] \approx -0.0147$: the default action lies to the left of all three agendas, every expert reveals only states above it, and the decision-maker is much better off.

So an enlarged panel can have several equilibria even when the initial panel has a unique one, and the new equilibria can also be worse, even when the equilibrium that continues the initial one improves.

Example 4. Consider a panel of two experts with $x_1 = 0.4$, $a_1 = 0.5$, $x_2 = 0.7$, $a_2 = 0.8$, again with a uniform prior and quadratic loss. It has a unique equilibrium, with $y^* \approx 0.3050$ and $\mathbb{E}[U] \approx -0.03161$; the default action lies to the left of both agendas, so both experts reveal only states above it. Add an expert with the extreme agenda $x_3 = 0$ and quality $a_3 = 0.5$, who reveals states below the default action. In the equilibrium that continues the initial one, $y^* \approx 0.3873$ and the decision-maker is better off, $\mathbb{E}[U] \approx -0.02502$. But the enlarged panel also has two new equilibria, in which the default action lies between the agendas of the initial experts: $y^* \approx 0.4297$ with $\mathbb{E}[U] \approx -0.03290$, and $y^* \approx 0.5586$ with $\mathbb{E}[U] \approx -0.04902$. Hence the panel is unstable with respect to this expert, although the equilibrium that continues the initial one improves.

This is why Definition 1 below requires the decision-maker to be better off in every equilibrium of the enlarged panel.[2]

[2]All equilibria in Examples 2–4 were computed exactly: with a uniform prior and quadratic loss, the left-hand side of (2) is a quadratic polynomial in y^* between consecutive breakpoints x_i, $2x_i$ and $2x_i - 1$.

Proposition BM (Bhattacharya and Mukherjee 2013, Proposition 3)**.** *In the game with opposite extreme agendas $x_1 = 0$, $x_2 = 1$ and any qualities $a_1, a_2 \in (0, 1)$, an increase in the quality of either expert increases the decision-maker's expected utility.*

This proposition was established in BM. Adding a new expert with agenda $x_3 = 0$ and quality a_3 is equivalent to increasing the quality of the first expert to $a_1' = 1-(1-a_1)(1-a_3)$. Hence the example of an improving equilibrium is trivial, and no numbers are needed here, since the sufficient conditions are already given by BM.

In the next few paragraphs I demonstrate that, despite the ambiguous direction of the change in the decision-maker's utility, one can identify special classes of panels for which the outcome can be determined, i.e., for which one can say what may and may not happen with the decision-maker's utility. Before doing so, I introduce the main definition of this paper, which justifies its title and clarifies what exactly I study.

Definition 1. Consider a panel of experts and an equilibrium of it with default action y_0. The panel is *stable* (at y_0) with respect to a new expert with agenda x' and quality a' if the expected utility of the decision-maker in every equilibrium of the enlarged panel is not lower than in the initial equilibrium. Otherwise the panel is *unstable* (at y_0) with respect to this expert. When the initial panel has a unique equilibrium, I omit the reference to y_0.

Note that stability depends on the new expert, and that the definition refers to one particular equilibrium of the initial panel: when the initial panel has multiple equilibria, I consider one of them and do not try to "match" different equilibria of the initial panel with those of the enlarged one, considering them in pairs. Before describing the classes of panels for which one can determine whether they are stable, I decompose the effect of a new expert into two parts.

3.2 Direct and strategic effects

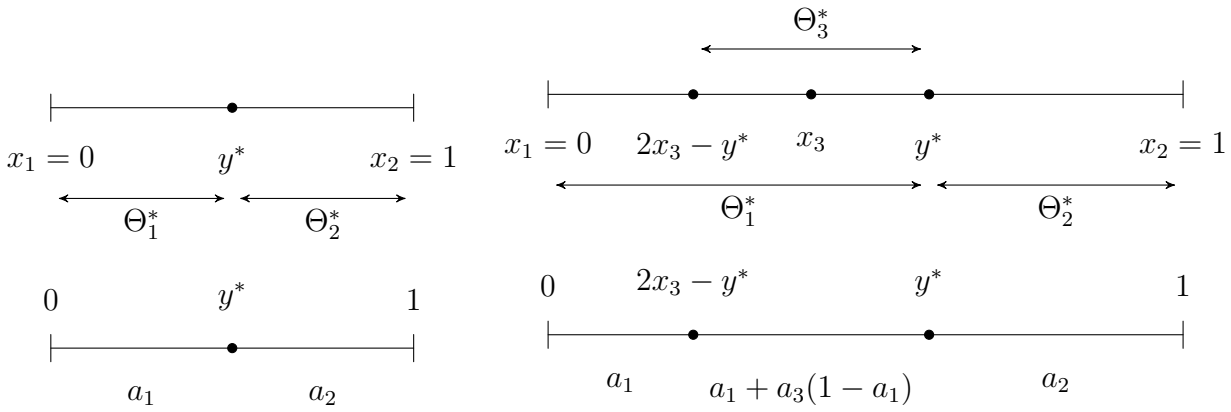
The change in the decision-maker's utility after the addition of a new expert is the result of two effects. Below I describe both effects and how they interact in the cases of extreme and non-extreme agendas.

1. The first effect is very simple. Imagine that after the addition of a new expert the default action changes only very slightly, so that this change in the equilibrium can be neglected. Then the revelation sets of the initial experts remain the same, and the only change in the model is the appearance of a new revelation set, which increases the probability that the decision-maker learns the true state of the world. This always increases the utility of the decision-maker; I call it the *direct effect*. Figure 1 illustrates this effect. The lower axis shows the probability that a state in each region is revealed to the decision-maker.

2. If we relax the assumption we made, a more complicated interaction appears. When the default action changes, the revelation sets of the initial experts change, and the direction of this effect is ambiguous. Indeed, consider the situation in which the default action lies between the two agendas and moves to the left. Then the left revelation set shrinks and the right one expands, and in general it is impossible to say which of these changes dominates. I call this effect of the change in the revelation sets the *strategic effect*. Figure 2 illustrates the strategic effect in the non-extreme agenda case, and Figure 3 can be seen as an illustration of the strategic effect in the extreme agenda case.

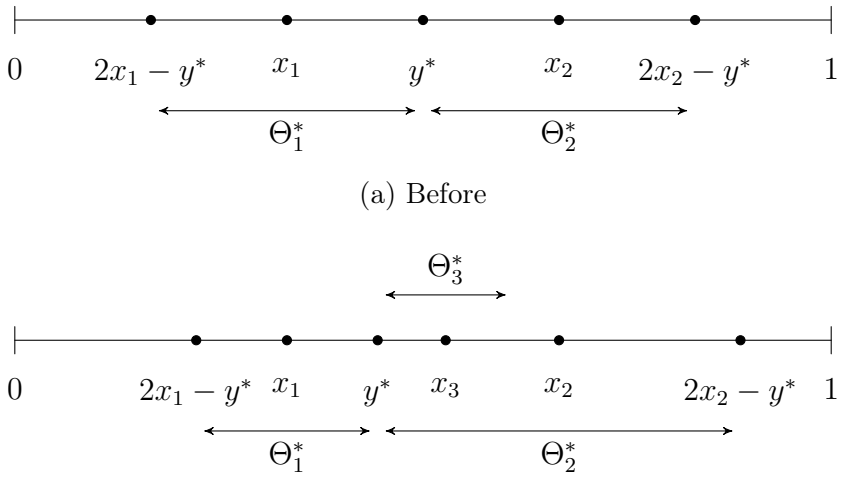
In the next paragraphs I describe the intuition behind the important difference between the behavior of the strategic effect in the non-extreme and in the extreme agenda case; the direct effect, as noted above, is always the same.

If the agendas are sufficiently extreme, the boundaries of the revelation sets coincide with the boundaries of the interval and with the default action, so the change in them



(a) Before (b) After, with the default action held fixed

Figure 1: The direct effect (lower axes: probability that a state in each region is revealed)



(a) Before

(b) After

Figure 2: The strategic effect

is benign: it only redistributes states between these sets in the region around the default action. Moreover, the appearance of a new revelation set, which increases the probability of revelation for some states of the world, gives an additional increase in the expected utility of the decision-maker.

When the agendas are insufficiently extreme, however, the outer boundaries of the revelation sets also change when the default action changes, and this amplifies the shrinking of one set and the expansion of the other. In Figure 2, as mentioned above, the revelation

set of the first expert shrinks and that of the second expert expands after the default action moves to the left. With extreme agendas, as in Figure 3, this effect is also present, but in a milder form. If the revelation set that shrinks is the more important one, this change may have a large negative effect on the utility of the decision-maker, large enough to outweigh the effect of the new revelation set described at the beginning of this subsection.

This is the intuition behind the complex interaction that takes place in the case of non-extreme experts, and it is what distinguishes this case from the case of extreme agendas. Sections 4 and 5 formalize it.

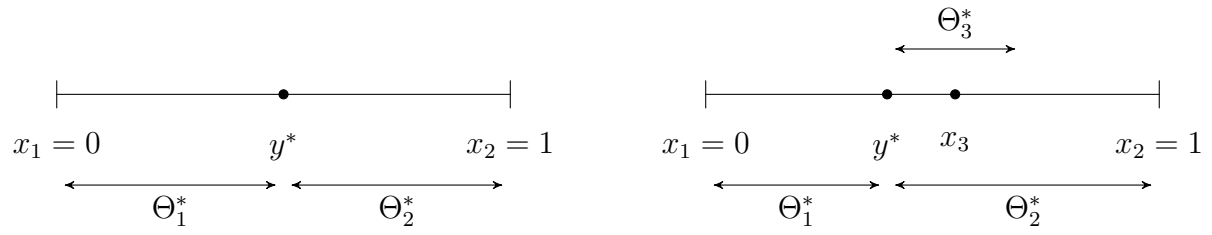
4 Extreme panels

Definition 2. A panel of experts is *extreme* if $x_i \in \{0, 1\}$ for all $i \in \{1, \ldots, n\}$.

Experts with the same agenda can be treated as a single expert: n experts with agenda 0 and qualities $a_1, \ldots, a_n$ act as one expert with quality $1 - \prod_{i=1}^{n}(1 - a_i)$, and similarly for agenda 1. An extreme panel is therefore, in effect, a panel of at most two experts with opposite agendas 0 and 1. Lemma 1 in the Appendix shows, by a local argument, that such a panel is stable with respect to any new expert of sufficiently low quality.

The main intuition behind this result is that the key factor determining the direction of the change in the decision-maker's utility is the change in the revelation sets of the experts in the initial panel. In the extreme case, the changes in these sets are rather limited, in the sense that they only redistribute states between the revelation sets: when y^* moves to the left or to the right, one revelation set expands and the other shrinks, but their outer boundaries, 0 and 1, do not move, wherever the new default action is. Figure 3 illustrates this stability of the revelation sets: the states far from the default action are always revealed with the same probability, whatever the default action.

As the results and examples below show, the main reason for a possible decrease in utility in the new equilibrium is the change in the boundaries of the revelation sets that are farthest



(a) Before (b) After

Figure 3: Revelation sets in an extreme panel before and after the addition of a new expert

from the default action. One boundary of each set is exactly the default action, and a change in it only "redistributes" the revelation sets around the default action; as shown in BM, in the case of extreme experts such a redistribution is always beneficial to the decision-maker when the quality of one of the experts increases. The situation after the appearance of a new revelation set is more complicated, but similar in this respect.

Moreover, these statements can be generalized if we recall how the revelation sets are constructed. Consider all default actions that can arise in equilibrium after the addition of a third expert, and let $[a, b]$ be the smallest interval containing them. If we replace the initial extreme agendas 0 and 1 with values from $[0, a/2]$ and $[(1 + b)/2, 1]$, respectively, the revelation sets of these experts do not change, because the boundaries of the state space $\Theta = [0, 1]$ are closer to the agendas than the default action is. These revelation sets remain $[0, y^*]$ and $[y^*, 1]$, respectively, so all the arguments remain valid.

In other words, this statement is close to the concept of "sufficiently extreme experts" in BM, who consider situations in which the revelation sets are $[0, y^*]$ or $[y^*, 1]$ and prove that if all revelation sets are of this form, then an increase in the quality of an expert increases the utility of the decision-maker. Let me formalize these ideas and draw the final conclusion of this subsection.

Definition 3. For a fixed initial panel of experts, the *future action set* Y^* is the set of all equilibrium default actions of the panels obtained by adding to it one expert with some agenda $x' \in [0, 1]$ and quality $a' \in [0, 1]$.

Definition 4. Let $[a, b]$ be the smallest interval containing the future action set Y^* of the initial panel. An expert of the initial panel is *sufficiently extreme* if his revelation set is $[0, y]$ for all $y \in [a, b]$ or $[y, 1]$ for all $y \in [a, b]$.

Remark 1. This is an implicit condition, but it can be reformulated more explicitly: an expert is sufficiently extreme if and only if his agenda satisfies $x_i \in [0, a/2] \cup [(1 + b)/2, 1]$.

The threshold on the quality of the new expert in Lemma 1 comes from the local nature of its proof. It can be equal to 1 for some new experts: if $x_1 = 0$, $x_2 = 1$ and $x_3 \in \{0, 1\}$, this follows trivially from Proposition BM. One may then ask whether there exists a panel for which the addition of any expert makes the decision-maker better off. In fact, every sufficiently extreme panel has this property.

Theorem 1. *(i) A panel of sufficiently extreme experts is stable with respect to any new expert, whatever his agenda $x' \in [0, 1]$ and quality $a' \in [0, 1]$.*

(ii) An extreme panel remains stable when any finite number of new experts is added at once: in every equilibrium of the enlarged panel the expected utility of the decision-maker is not lower than in the initial equilibrium.

Theorem 1 holds for any prior and any loss function satisfying the assumptions of Section 2; in particular, the balanced panel with $x_1 = 0$, $x_2 = 1$ and $a_1 = a_2$ is stable with respect to any new expert. The reason is that in a sufficiently extreme panel the only boundary of an initial expert's revelation set that moves with the default action is the default action itself, where the loss of the decision-maker is zero. Hence, if the revelation set of the new expert is held fixed at its equilibrium value, the equilibrium default action maximizes the decision-maker's payoff among all default actions she could commit to, including the initial one; and at the initial default action the new expert can only add disclosure.

5 Non-extreme panels

5.1 Instability

If instead the agendas of the initial experts are far from the boundaries of $[0,1]$, the utility of the decision-maker can decrease after the addition of a new expert: when the outer boundaries of the revelation sets move with the default action, the strategic effect described in Section 3.2 can outweigh the direct one. The next theorem formalizes this.

Theorem 2. *Consider a panel of two experts and an equilibrium of it with default action y_0 such that $x_1 < y_0 < x_2$ and neither expert is sufficiently extreme at y_0, i.e., $2x_1 - y_0 > 0$ and $2x_2 - y_0 < 1$. Suppose that $T \neq 0$, where*

$$T = a_1 v(2(y_0 - x_1)) f(2x_1 - y_0) - a_2 v(2(x_2 - y_0)) f(2x_2 - y_0). \tag{T}$$

Then for every quality $a_3 \in (0,1)$ there is an agenda x_3 such that the panel is unstable at y_0 with respect to the expert with agenda x_3 and quality a_3.

The quantity T is the derivative of the decision-maker's expected utility with respect to the default action at y_0 when the revelation sets respond to it: a small shift of the default action affects her utility only through the outer boundaries of the revelation sets. The condition $T \neq 0$ cannot be dropped: when $T = 0$, the strategic effect of a new expert vanishes to first order in his quality.

Proposition 1. *Consider a panel of two experts and an equilibrium of it with default action y_0 as in Theorem 2, but with $T = 0$. Suppose that y_0 is the unique equilibrium of the panel and that $V'(y_0) \neq 0$. Then for every agenda x' there is a threshold $\overline{a}(x') > 0$ such that the panel is stable with respect to the expert with agenda x' and any quality $a' \leq \overline{a}(x')$.*

Theorem 2 and Proposition 1 show that, for the panels they consider, the sign of the strategic effect of a new expert of low quality is governed by T: if $T \neq 0$, some expert is

harmful, and if $T = 0$, no expert of sufficiently low quality is. In an extreme panel $T = 0$ automatically, because the outer boundaries of the revelation sets do not move. Among non-extreme panels, $T = 0$ holds, for instance, for every symmetric panel ($x_1 + x_2 = 1$, $a_1 = a_2$) with a prior symmetric around $1/2$, where $y_0 = 1/2$ and the two terms of T cancel; in the numerical examples of Appendix A.5, such panels are stable with respect to any single additional expert, of any agenda and quality.

5.2 Which experts are harmful?

Theorem 2 shows that harmful experts exist, and its proof constructs them close to the default action. When the decision-maker chooses experts from a given pool, however, the relevant question is which experts are harmful. The next proposition answers it in terms of two functions of the initial panel alone. The first is the left-hand side $V_0(y)$ of the equilibrium condition (2). The second is

$$U(y) = \int_0^1 P_0(\theta; y)\, u_J(y, \theta)\, dF(\theta),$$

where P_0 is the probability of nondisclosure in the initial panel: $U(y)$ is the decision-maker's payoff when the default action is y and the experts best respond to it, i.e., the payoff she would obtain by committing to y. At an equilibrium y_0 of the initial panel, $\mathbb{E}[U_0] = U(y_0)$.

Proposition 2. *Let y_0 be an equilibrium of a panel, and let y_1 be an equilibrium of the panel enlarged by an expert with agenda $x' \neq y_1$ and quality a'. Then*

$$\mathbb{E}[U_1] = U(y_1) + D, \qquad D = a' \int_{\Theta'(y_1)} P_0(\theta; y_1)\, v(|y_1 - \theta|)\, dF(\theta),$$

and $0 \leq D \leq 2|x' - y_1|\,|V_0(y_1)|$. Consequently:

(i) if the decision-maker is worse off in the equilibrium y_1 than at y_0, then $U(y_1) < U(y_0)$;

(ii) if $U(y_0) - U(y_1) > 2|x' - y_1|\,|V_0(y_1)|$, then she is worse off in the equilibrium y_1, and

the panel is unstable at y_0 with respect to this expert.

D is the direct effect of the new expert at the default action y_1, and $U(y_1) - U(y_0)$ is the strategic effect. Part (i) says that a harmful expert must move the default action into the region where the commitment payoff of the initial panel is lower than at y_0; near y_0, this is the direction opposite to the sign of T in Theorem 2. Part (ii) identifies the harmful experts as narrow but strong ones. The revelation set of the new expert has y_1 as an endpoint and length at most $2|x' - y_1|$, so an expert whose agenda is close to the default action he induces discloses little; but the equilibrium condition requires his disclosure to move the default action by exactly as much as $|V_0(y_1)|$ demands, which a sufficiently informed expert can do however narrow his revelation set is. The necessary condition (i) is far from sufficient: in an extreme panel U is maximized at y_0 (Section 6), so every $y_1 \neq y_0$ satisfies it, and yet, by Theorem 1, no expert is harmful.

Example 5. In the panel of Example 2, $U(y) < U(y_0)$ exactly for $y > y_0 \approx 0.4711$. By Proposition 2(i), a new expert can therefore harm the decision-maker only by moving the default action to the right, towards the better informed expert, whose outer boundary then moves inward. The expert of Example 2 does exactly that: with $x_3 = 0.45$ and $a_3 = 0.5$ the default action moves to $y_1 \approx 0.4720$, and $U(y_0) - U(y_1) \approx 4.9 \cdot 10^{-5}$ exceeds the bound $2|x_3 - y_1||V_0(y_1)| \approx 3.8 \cdot 10^{-5}$ on his direct effect, so condition (ii) holds.

6 Commitment and the fragility of moderate panels

The results above may seem to contradict practice, because many policymakers would like to have representatives of the whole "political spectrum" in their advisory committees and would not like to rely only on the opinions of radical experts who take extreme left or right positions. But according to the results above, such radical committees are in fact much better for the decision-maker, because a new expert can be added to them with no doubt about the benefits that he brings.

In the following paragraphs, I describe in simple terms the problem that may arise within a panel that is not sufficiently extreme, in order to demonstrate why some assumptions of the original BM paper are important for the results of this paper. With no information acquisition costs and verifiable messages, an expert is left with only one strategic instrument of manipulation: he can conceal information about the true state of the world. The decision-maker takes this into account, and she knows that when the experts together say nothing about the state of the world, it may be either that they do not know anything or that the state does not belong to their revelation sets.

It might seem that with this knowledge the decision-maker cannot be worse off, but in fact she has no commitment power over her default action, and hence her future choice of action influences the revelation sets of the experts. If the decision-maker had commitment power, there would be no problem of a worsening equilibrium. This is true because she could keep the status quo default action, and the strategic effect would not appear. This observation is very simple and the reasoning is clear, so I state it formally.

Proposition 3. *When the decision-maker has commitment power over the default action, every panel of experts is stable with respect to any new expert, for all* $x_i \in [0,1]$ *and* $a_i \in (0,1)$, $i \in \{1,\dots,n\}$.

Proposition 3 also gives a compact way to state the difference between the two classes of panels: extreme panels behave as if the decision-maker could commit to her default action. In an extreme panel, the only boundary of a revelation set that moves with the default action is the default action itself, where the loss of the decision-maker is zero. Hence the derivative of her expected utility with respect to the default action, with the revelation sets responding to it, equals the left-hand side of (2), and the equilibrium default action is exactly the one she would commit to. In the two-expert panel of Theorem 2, which is not sufficiently extreme, the outer boundaries of the revelation sets move as well, and at the equilibrium this derivative equals T in (T); whenever $T \neq 0$, she would like to commit to a different default action. To first order in its quality, a new expert can harm her only by moving the default

action in the direction opposite to the sign of T, and Theorem 2 shows that an expert whose agenda is close enough to the default action, on the appropriate side, does exactly that. In Example 2, $T < 0$: the decision-maker would like to move the default action to the left, towards the less informed expert, which would extend the disclosure of the better-informed expert at the right end of the spectrum. The new expert at $x_3 = 0.45$ instead reveals states just below the default action and pushes it to the right, so the better-informed expert stops reporting some of the extreme states on his side.

When commitment is not possible, the default action changes after the appearance of a new expert as the result of strategic interaction. When the initial experts are not sufficiently extreme, their revelation sets may shrink, so, in simple words, they may provide less information about the very extreme states of the world, which they particularly dislike. So the main problem of a non-extreme panel is that an increase in the probability of revealing the states in the middle of the spectrum leads to a decrease in the probability of revealing the extreme states of the world. This can never happen in an extreme panel, because such experts always prefer to reveal these extreme states, which are closer to their agendas than the default action. Basically, the findings of the whole paper can be summarized by the idea that the panel of experts is stable unless, after the addition of a new expert, the experts start to conceal information about some extreme states of the world.

Finally, the local argument of Lemma 1 compares the direct and the strategic effects for a new expert of low quality, and for such an expert the strategic effect in an extreme panel has the same sign as the direct one. For a new expert of high quality the enlarged panel may have several equilibria, and the local argument no longer applies, but Theorem 1 shows that in a sufficiently extreme panel the strategic effect can never outweigh the direct one, because, as explained above, the equilibrium default action is always the one the decision-maker would commit to.

References

Austen-Smith, D. (1990). Information transmission in debate. *American Journal of Political Science 34*(1), 124–152.

Austen-Smith, D. (1993). Interested experts and policy advice: Multiple referrals under open rule. *Games and Economic Behavior 5*(1), 3–43.

Battaglini, M. (2002). Multiple referrals and multidimensional cheap talk. *Econometrica 70*(4), 1379–1401.

Beniers, K. J. and O. H. Swank (2004). On the composition of committees. *The Journal of Law, Economics, and Organization 20*(2), 353–378.

Bhattacharya, S., M. Goltsman, and A. Mukherjee (2018). On the optimality of diverse expert panels in persuasion games. *Games and Economic Behavior 107*, 345–363.

Bhattacharya, S. and A. Mukherjee (2013). Strategic information revelation when experts compete to influence. *The RAND Journal of Economics 44*(3), 522–544.

Cai, H. (2009). Costly participation and heterogeneous preferences in informational committees. *The RAND Journal of Economics 40*(1), 173–189.

Dewatripont, M. and J. Tirole (1999). Advocates. *Journal of Political Economy 107*(1), 1–39.

Dye, R. A. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research 23*(1), 123–145.

Gentzkow, M. and E. Kamenica (2017). Competition in persuasion. *The Review of Economic Studies 84*(1), 300–322.

Gerardi, D. and L. Yariv (2008). Costly expertise. *American Economic Review 98*(2), 187–193.

Gilligan, T. W. and K. Krehbiel (1989). Asymmetric information and legislative rules with a heterogeneous committee. *American Journal of Political Science 33*(2), 459–490.

Glazer, J. and A. Rubinstein (2004). On optimal rules of persuasion. *Econometrica 72*(6), 1715–1736.

Glazer, J. and A. Rubinstein (2006). A study in the pragmatics of persuasion: A game theoretical approach. *Theoretical Economics 1*(4), 395–410.

Gong, Q. and H. Yang (2018). Balance of opinions in expert panels. *Economics Letters 170*, 151–154.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics 24*(3), 461–483.

Jung, W.-O. and Y. K. Kwon (1988). Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research 26*(1), 146–153.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Kartik, N., F. X. Lee, and W. Suen (2017). Investment in concealable information by biased experts. *The RAND Journal of Economics 48*(1), 24–43.

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

Milgrom, P. and J. Roberts (1986). Relying on the information of interested parties. *The RAND Journal of Economics 17*(1), 18–32.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics 12*(2), 380–391.

Okuno-Fujiwara, M., A. Postlewaite, and K. Suzumura (1990). Strategic information revelation. *The Review of Economic Studies 57*(1), 25–47.

Shavell, S. (1994). Acquisition and disclosure of information prior to sale. *The RAND Journal of Economics 25*(1), 20–36.

Shin, H. S. (1994). The burden of proof in a game of persuasion. *Journal of Economic Theory 64*(1), 253–264.

Shin, H. S. (1998). Adversarial and inquisitorial procedures in arbitration. *The RAND Journal of Economics 29*(2), 378–405.

A Proofs

Throughout the Appendix, V denotes the left-hand side of the equilibrium condition (2) for the panel under consideration, and $\mathbb{E}[U]$ the expected utility (3) of the decision-maker in the equilibrium under consideration. Derivatives of $\mathbb{E}[U]$ and y^* with respect to a_3 are taken along a branch of equilibria. When $\mathbb{E}[U]$ is differentiated with respect to a_3, the terms that come from the dependence of the integrands on y^* add up to $V \cdot \frac{dy^*}{da_3} = 0$, and the boundary terms at $\theta = y^*$ vanish because $u_J(y^*, y^*) = 0$; only the boundary terms at the other ends of the revelation sets remain.

A.1 Local stability of sufficiently extreme panels

Lemma 1. *For a panel of sufficiently extreme experts and a new expert with agenda x', there exists a threshold $\overline{a}(x') > 0$ such that, whenever $a' \leq \overline{a}(x')$, the enlarged panel has a unique equilibrium and the panel is stable with respect to the new expert. In particular, this holds for a panel of n identical extreme experts and for a panel of two experts with opposite extreme agendas.*

Proof. I first prove the lemma for two experts with opposite extreme agendas, $x_1 = 0$ and $x_2 = 1$, and a new expert with agenda x_3 and quality a_3, then for identical extreme experts, and finally for sufficiently extreme experts.

Opposite extreme experts. The initial panel has a unique equilibrium y_0 by the argument of Example 1: the left-hand side of (2) is strictly decreasing, because the boundary terms vanish. Depending on the position of the default action y relative to x_3, the revelation set of the new expert takes one of four forms (not all of them need to be possible, depending on the value of x_3):

1. $y \geq 2x_3$: then $\Theta_3 = [0, y]$, the same as the revelation set of the first expert. The enlarged panel is then equivalent to the initial one with the quality of the first expert increased to $1 - (1 - a_1)(1 - a_3)$, and, according to Proposition BM, the utility of the decision-maker increases.

2. $x_3 \leq y \leq 2x_3$: then $\Theta_3 = [2x_3 - y, y]$.

3. $2x_3 - 1 \leq y \leq x_3$: then $\Theta_3 = [y, 2x_3 - y]$.

4. $y \leq 2x_3 - 1$: then $\Theta_3 = [y, 1]$, which is equivalent to an increase in the quality of the second expert, and, according to Proposition BM, the utility of the decision-maker increases.

Uniqueness for small a_3. In each of the four regions, $\partial V/\partial y$ is the sum of the integral $\int_0^1 P(\theta; y) u_J''(y, \theta) dF(\theta)$, which is negative and, for $a_3 \leq 1/2$, bounded away from zero uniformly in y (since $P \geq (1 - a_1)(1 - a_2)/2$), and of at most one boundary term, which is proportional to a_3 (see cases 2 and 3 below). Hence there exists $\overline{a}_3(x_3) > 0$ such that $\partial V/\partial y < 0$ for all y and all $a_3 \leq \overline{a}_3(x_3)$. For such a_3, V is strictly decreasing in y, so the enlarged panel has a unique equilibrium $y^*(a_3)$, which is continuous in a_3, with $y^*(0) = y_0$. It remains to show that $\mathbb{E}[U]$ is nondecreasing in a_3 along this equilibrium. In cases 1 and 4

this follows from Proposition BM; I prove it for case 2 and then for case 3, whose proof is the same in technique.

Case 2. Suppose that $y^* \in [x_3, \min(2x_3, 1)]$. If $y^* = x_3$, then $\Theta_3 = \{y^*\}$ has measure zero, so the new expert does not change the equilibrium or the expected utility; hence assume $y^* > x_3$. The equilibrium condition and the utility of the decision-maker are

$$V = (1-a_1)\int_0^{2x_3-y^*} u_J'(y^*,\theta)dF + (1-a_1)(1-a_3)\int_{2x_3-y^*}^{y^*} u_J'(y^*,\theta)dF + (1-a_2)\int_{y^*}^1 u_J'(y^*,\theta)dF = 0,$$

$$\mathbb{E}[U] = (1-a_1)\int_0^{2x_3-y^*} u_J(y^*,\theta)dF + (1-a_1)(1-a_3)\int_{2x_3-y^*}^{y^*} u_J(y^*,\theta)dF + (1-a_2)\int_{y^*}^1 u_J(y^*,\theta)dF.$$

The derivative with respect to a_3 is

$$\frac{d\mathbb{E}[U]}{da_3} = -(1-a_1)\int_{2x_3-y^*}^{y^*} u_J(y^*,\theta)dF + \frac{dy^*}{da_3}\left[-a_3(1-a_1)u_J(y^*,2x_3-y^*)f(2x_3-y^*)\right].$$

Note first that $-(1-a_1)\int_{2x_3-y^*}^{y^*} u_J(y^*,\theta)dF > 0$ and $-a_3(1-a_1)u_J(y^*,2x_3-y^*)f(2x_3-y^*) > 0$. By the implicit function theorem, $\frac{dy^*}{da_3} = -\frac{\partial V/\partial a_3}{\partial V/\partial y^*}$, where

$$\frac{\partial V}{\partial a_3} = -(1-a_1)\int_{2x_3-y^*}^{y^*} u_J'(y^*,\theta)dF > 0,$$

$$\frac{\partial V}{\partial y^*} = -a_3(1-a_1)u_J'(y^*,2x_3-y^*)f(2x_3-y^*) + (1-a_1)\int_0^{2x_3-y^*} u_J''(y^*,\theta)dF + (1-a_1)(1-a_3)\int_{2x_3-y^*}^{y^*} u_J''(y^*,\theta)dF + (1-a_2)\int_{y^*}^1 u_J''(y^*,\theta)dF.$$

The first term is positive, while the three integrals are negative; since $a_3 \leq \overline{a}_3(x_3)$, $\frac{\partial V}{\partial y^*} < 0$. Hence $\frac{dy^*}{da_3} > 0$, and $\frac{d\mathbb{E}[U]}{da_3} > 0$.

Case 3. Suppose that $y^* \in [\max(2x_3 - 1, 0), x_3]$. As in case 2, if $x_3 = y^*$ nothing changes, so assume $y^* < x_3$. The equilibrium condition and the utility of the decision-maker are

$$V = (1 - a_1) \int_0^{y^*} u'_J(y^*, \theta) dF + (1 - a_2)(1 - a_3) \int_{y^*}^{2x_3 - y^*} u'_J(y^*, \theta) dF + (1 - a_2) \int_{2x_3 - y^*}^{1} u'_J(y^*, \theta) dF = 0,$$

$$\mathbb{E}[U] = (1 - a_1) \int_0^{y^*} u_J(y^*, \theta) dF + (1 - a_2)(1 - a_3) \int_{y^*}^{2x_3 - y^*} u_J(y^*, \theta) dF + (1 - a_2) \int_{2x_3 - y^*}^{1} u_J(y^*, \theta) dF.$$

The derivative with respect to a_3 is

$$\frac{d\mathbb{E}[U]}{da_3} = -(1 - a_2) \int_{y^*}^{2x_3 - y^*} u_J(y^*, \theta) dF + \frac{dy^*}{da_3} \left[a_3(1 - a_2) u_J(y^*, 2x_3 - y^*) f(2x_3 - y^*) \right].$$

Note first that $-(1-a_2) \int_{y^*}^{2x_3 - y^*} u_J(y^*, \theta) dF > 0$ and $a_3(1-a_2) u_J(y^*, 2x_3 - y^*) f(2x_3 - y^*) < 0$. As in case 2,

$$\frac{\partial V}{\partial a_3} = -(1 - a_2) \int_{y^*}^{2x_3 - y^*} u'_J(y^*, \theta) dF < 0,$$

$$\frac{\partial V}{\partial y^*} = a_3(1 - a_2) u'_J(y^*, 2x_3 - y^*) f(2x_3 - y^*) + (1 - a_1) \int_0^{y^*} u''_J(y^*, \theta) dF + (1 - a_2)(1 - a_3) \int_{y^*}^{2x_3 - y^*} u''_J(y^*, \theta) dF + (1 - a_2) \int_{2x_3 - y^*}^{1} u''_J(y^*, \theta) dF.$$

Again, the first term is positive and the integrals are negative, so for $a_3 \leq \overline{a}_3(x_3)$ we have $\frac{\partial V}{\partial y^*} < 0$, hence $\frac{dy^*}{da_3} < 0$ and $\frac{d\mathbb{E}[U]}{da_3} > 0$.

Conclusion. In all four cases $\mathbb{E}[U]$ is continuous in a_3 on $[0, \overline{a}_3(x_3)]$ and its derivative is nonnegative wherever it exists, so $\mathbb{E}[U](a_3) \geq \mathbb{E}[U](0)$ for all $a_3 \leq \overline{a}_3(x_3)$. As we can easily see, $\mathbb{E}[U](0)$ is the initial utility of the decision-maker, and for such a_3 the enlarged panel has a unique equilibrium, so the panel is stable.

Identical extreme experts. Consider the proof above with $a_2 = 0$, and let $a_1 = 1 - \prod_{i=1}^{n}(1 - a_i)$ be the quality of the single expert that replaces the group of experts located at $x = 0$. Then the proof is exactly the one we need; in cases 1 and 4, Proposition BM extends to $a_2 = 0$ by continuity of the equilibrium utility in a_2 (the equilibrium is unique and continuous in a_2).

Sufficiently extreme experts. The two proofs above never use the fact that $x_i = 0$ or $x_i = 1$, only the revelation sets of these experts. By Definition 4, when the default action moves from y_0 to y^*, the revelation sets remain $[0, y^*]$ or $[y^*, 1]$, because either $x_i \leq a/2 \leq y^*/2$ or $x_i \geq (1 + b)/2 \geq (1 + y^*)/2$. (The idea behind the definition of sufficiently extreme experts is exactly that their revelation sets remain $[0, y^*]$ or $[y^*, 1]$.) So the proofs remain the same, and hence for every initial panel of sufficiently extreme experts there exists a threshold as in the two cases above. □

A.2 Proof of Theorem 1

Part (i). Let $[a, b]$ be the smallest interval containing the future action set Y^* of the initial panel. By Definition 4, for every $y \in [a, b]$ each initial expert has revelation set $[0, y]$ or $[y, 1]$, so that

$$P_0(\theta; y) = \alpha\, \mathbf{1}\{\theta < y\} + \beta\, \mathbf{1}\{\theta > y\}, \qquad \alpha = \prod_{i:\, \Theta_i(y) = [0,y]} (1 - a_i), \quad \beta = \prod_{i:\, \Theta_i(y) = [y,1]} (1 - a_i),$$

where P_0 is the probability of nondisclosure in the initial panel. In particular, on $[a, b]$ the left-hand side of (2) for the initial panel has derivative $\int_0^1 P_0(\theta; y) u_J''(y, \theta) dF < 0$ (the

boundary terms at $\theta = y$ vanish because $u'_J(y, y) = 0$), so the initial panel has a unique equilibrium y_0, which lies in $[a, b]$ since $y_0 \in Y^*$ (take $a' = 0$).

Let y^* be any equilibrium of the enlarged panel; $y^* \in Y^* \subseteq [a, b]$. Hold the revelation set of the new expert fixed at its equilibrium value, and let $q(\theta) = 1 - a'\mathbf{1}\{\theta \in \Theta'(y^*)\} \in [0, 1]$. Define, for $y \in [a, b]$,

$$G(y) = \int_0^1 q(\theta) P_0(\theta; y) u_J(y, \theta) dF(\theta) = \alpha \int_0^y q\, u_J(y, \theta) dF + \beta \int_y^1 q\, u_J(y, \theta) dF,$$

the payoff of the decision-maker if the default action is y, the initial experts best respond to y and the new expert keeps the revelation set $\Theta'(y^*)$. Since $u_J(y, y) = u'_J(y, y) = 0$, the boundary terms at $\theta = y$ vanish in both derivatives, and

$$G'(y) = \int_0^1 q(\theta) P_0(\theta; y) u'_J(y, \theta) dF, \qquad G''(y) = \int_0^1 q(\theta) P_0(\theta; y) u''_J(y, \theta) dF \leq 0.$$

If $qP_0 = 0$ almost everywhere, every state is disclosed in the enlarged panel, $\mathbb{E}[U^*] = 0$, and there is nothing to prove. Otherwise $G'' < 0$, so G is strictly concave on $[a, b]$. Since the probability of nondisclosure in the enlarged panel at the default action y^* is $q(\theta)P_0(\theta; y^*)$, we have $G'(y^*) = 0$ by (2); hence y^* maximizes G on $[a, b]$, and

$$\mathbb{E}[U^*] = G(y^*) \geq G(y_0) = \int_0^1 q(\theta) P_0(\theta; y_0) u_J(y_0, \theta) dF \geq \int_0^1 P_0(\theta; y_0) u_J(y_0, \theta) dF = \mathbb{E}[U_0],$$

where the last inequality holds because $q \leq 1$ and $u_J \leq 0$.

Part (ii). If $x_i \in \{0, 1\}$ for all initial experts, their revelation sets are $[0, y]$ or $[y, 1]$ for every $y \in [0, 1]$, so the argument of part (i) applies on the whole interval $[0, 1]$, with $q(\theta) = \prod_k \left(1 - a'_k \mathbf{1}\{\theta \in \Theta'_k(y^*)\}\right)$ for the new experts k. □

A.3 Proof of Theorem 2

Let N be an open interval around y_0 such that $x_1 < y < x_2$, $2x_1 - y > 0$ and $2x_2 - y < 1$ for all $y \in N$. For $y \in N$ the revelation sets of the initial experts are $\Theta_1(y) = [2x_1 - y, y]$ and $\Theta_2(y) = [y, 2x_2 - y]$. Denote by $V_0(y)$ the left-hand side of (2) for the initial panel, and let

$$U(y) = \int_0^{2x_1-y} u_J(y,\theta)dF + (1-a_1)\int_{2x_1-y}^{y} u_J(y,\theta)dF \\ + (1-a_2)\int_y^{2x_2-y} u_J(y,\theta)dF + \int_{2x_2-y}^{1} u_J(y,\theta)dF$$

be the expected utility of the decision-maker when the default action is y and the two experts best respond to it, i.e., the payoff she would get by committing to y. Her equilibrium utility is $\mathbb{E}[U_0] = U(y_0)$.

Step 1: the derivative of U. When U is differentiated, the terms coming from the integrands add up to $V_0(y)$, the boundary terms at $\theta = y$ vanish because $u_J(y,y) = 0$, and the outer boundaries $2x_1 - y$ and $2x_2 - y$, which move with derivative -1, give

$$U'(y) = V_0(y) + B(y), \qquad B(y) = -a_1 u_J(y, 2x_1 - y) f(2x_1 - y) + a_2 u_J(y, 2x_2 - y) f(2x_2 - y).$$

B is continuous and, since $u_J(y,\theta) = -v(|y-\theta|)$, $B(y_0) = T$ as defined in (T). As $V_0(y_0) = 0$, we get $U'(y_0) = T$. Moreover, V_0 is Lipschitz on N: its derivative is $\int_0^1 P(\theta; y) u_J''(y,\theta)dF$ plus two boundary terms at the outer boundaries, and all of them are bounded, since $\left|\int_0^1 u_J''(y,\theta)dF\right| \le 2 \max f \cdot v'(1)$. Let K be a Lipschitz constant of V_0 on N.

Step 2: the target default action. Let $\sigma = \operatorname{sign}(T)$ and $y_1 = y_0 - \sigma\varepsilon$ for $\varepsilon > 0$. Since U' is continuous and $U'(y_0) = T \neq 0$, for all sufficiently small ε the mean value theorem gives

$$U(y_0) - U(y_1) \ge \frac{|T|}{2}\,\varepsilon,$$

and $|V_0(y_1)| \leq K\varepsilon$ because $V_0(y_0) = 0$. If $V_0(y_1) = 0$, then y_1 is itself an equilibrium of the initial panel with a lower utility than y_0; adding an expert with agenda $x_3 = y_1$ and any quality, whose revelation set at the default action y_1 is $\{y_1\}$ and has measure zero, leaves y_1 an equilibrium of the enlarged panel, so the panel is unstable at y_0. Hence assume $V_0(y_1) \neq 0$, and suppose first that $V_0(y_1) < 0$.

Step 3: the new expert. Fix $a_3 \in (0,1)$. For $d \in (0, y_1 - x_1]$, place the new expert at $x_3 = y_1 - d$. At the default action y_1 his revelation set is $\Theta_3(y_1) = [y_1 - 2d, y_1]$, which lies in $\Theta_1(y_1)$, where the initial probability of nondisclosure is $1 - a_1$. Since $u'_J(y_1, \theta) = -v'(y_1 - \theta)$ for $\theta < y_1$, the left-hand side of (2) for the enlarged panel at the default action y_1 equals

$$V_0(y_1) + a_3(1 - a_1)\,\phi(d), \qquad \phi(d) = \int_{y_1 - 2d}^{y_1} v'(y_1 - \theta) dF(\theta).$$

The function ϕ is continuous and strictly increasing, $\phi(0) = 0$, and $\phi(d) \geq \min f \cdot v(2d)$. For ε small enough, $|V_0(y_1)| \leq K\varepsilon < a_3(1-a_1) \min f \cdot v(y_0 - x_1) \leq a_3(1-a_1)\phi(y_1 - x_1)$, so there exists $d = d(\varepsilon) \in (0, y_1 - x_1]$ with $a_3(1 - a_1)\phi(d) = -V_0(y_1)$. Moreover, $d(\varepsilon) \to 0$ as $\varepsilon \to 0$, because $v(2d(\varepsilon)) \leq K\varepsilon/(a_3(1 - a_1) \min f)$.

With this choice, y_1 is an equilibrium default action of the enlarged panel: given the default action y_1, the three experts' revelation sets are $\Theta_1(y_1)$, $\Theta_2(y_1)$ and $\Theta_3(y_1)$, and for these sets y_1 satisfies (2), which, by strict concavity of u_J in y, makes it the decision-maker's best response.

Step 4: the utility of the decision-maker. In this equilibrium the decision-maker's utility is

$$\mathbb{E}[U_1] = U(y_1) + a_3(1 - a_1) \int_{y_1 - 2d}^{y_1} v(y_1 - \theta) dF(\theta),$$

where the second term is the direct effect of the new expert. Consider $g(x) = v'(x)x - v(x)$. Its derivative is $v''(x)x \geq 0$ and $g(0) = 0$; hence $v(x) \leq xv'(x)$ for all $x \geq 0$, and in particular $v(y_1 - \theta) \leq 2d\, v'(y_1 - \theta)$ for $\theta \in [y_1 - 2d, y_1]$. Therefore the direct effect is at most

$2d \cdot a_3(1-a_1)\phi(d) = 2d|V_0(y_1)| \leq 2K\varepsilon\, d(\varepsilon)$, and

$$\mathbb{E}[U_1] \leq U(y_0) - \varepsilon\left(\frac{|T|}{2} - 2K\, d(\varepsilon)\right) < U(y_0) = \mathbb{E}[U_0]$$

for ε small enough, since $d(\varepsilon) \to 0$.

If $V_0(y_1) > 0$, the argument is the same with the new expert at $x_3 = y_1 + d$, $d \in (0, x_2 - y_1]$, whose revelation set $[y_1, y_1 + 2d]$ lies in $\Theta_2(y_1)$, and with $1 - a_2$ in place of $1 - a_1$. Hence for every $a_3 \in (0,1)$ there is an agenda x_3 such that the enlarged panel has an equilibrium in which the decision-maker is strictly worse off than at y_0, i.e., the panel is unstable at y_0. □

Remarks. (i) The harmful expert moves the default action from y_0 to $y_0 - \sigma\varepsilon$, i.e., in the direction opposite to the one in which the decision-maker would like to move it if she could commit, and the direct benefit of his disclosure, of order $\varepsilon\, d(\varepsilon)$, is dominated by the strategic loss, of order ε.

(ii) The proof uses the two experts' outer boundaries only through B, and the only condition it needs is $T \neq 0$. If only one of the experts is not sufficiently extreme at y_0, say $2x_1 - y_0 < 0 < 2x_2 - y_0 < 1$, then the outer boundary of the first expert is fixed at 0 near y_0, B reduces to its second term, and $T = -a_2 v(2(x_2 - y_0)) f(2x_2 - y_0) < 0$; hence the conclusion of Theorem 2 holds in this case without any further condition.

(iii) When $W = V_0'(y_0) \neq 0$, the harmful equilibrium can also be found on the branch of equilibria through y_0. By the implicit function theorem, for a new expert of low quality the equilibrium moves in the direction of sign $\big(W(x_3 - y_0)\big)$, and at $a_3 = 0$

$$\frac{d\mathbb{E}[U]}{da_3} = (\text{direct effect}) + T\,\frac{dy^*}{da_3},$$

so the utility falls when x_3 is sufficiently close to y_0 and lies above y_0 if $WT < 0$ and below y_0 if $WT > 0$. In Example 2, $W \approx -1.015$ and $T \approx -0.058$, so $WT > 0$ and the harmful expert lies to the left of $y_0 \approx 0.471$, as $x_3 = 0.45$ does. The construction above does not

need $W \neq 0$: if $W = 0$ the branch through y_0 need not be differentiable, but the harmful equilibrium is constructed directly.

A.4 Proof of Proposition 1

Suppose $T = 0$, $W = V_0'(y_0) \neq 0$, and y_0 is the unique equilibrium of the initial panel, and consider a new expert with agenda $x_3 \neq y_0$ and quality a_3. (If $x_3 = y_0$, his revelation set at y_0 has measure zero, y_0 remains an equilibrium, and the argument below applies to the other equilibria.) Denote by $V(y; a_3)$ the left-hand side of (2) for the enlarged panel. As $a_3 \to 0$, $V(\cdot; a_3) \to V_0$ uniformly on $[0, 1]$, and near y_0 also $\partial V/\partial y \to V_0'$ uniformly, since the new expert changes the integrand by a term of order a_3 and adds at most one boundary term, which is proportional to a_3. Because V_0 has the single zero y_0 and $V_0'(y_0) \neq 0$, it follows that for a_3 small enough the enlarged panel has a unique equilibrium $y^*(a_3)$, which converges to y_0 and, by the implicit function theorem, is continuously differentiable in a_3 (for y near y_0 the position of $\Theta_3(y)$ relative to y does not change, since $x_3 \neq y_0$). Differentiating the utility along this equilibrium as in the proof of Theorem 2 gives, at $a_3 = 0$,

$$\frac{d\mathbb{E}[U]}{da_3} = -\int_{\Theta_3(y_0)} P(\theta; y_0) u_J(y_0, \theta) dF(\theta) + T \frac{dy^*}{da_3} = -\int_{\Theta_3(y_0)} P(\theta; y_0) u_J(y_0, \theta) dF(\theta) > 0,$$

since $T = 0$, $\Theta_3(y_0)$ has positive length and $f > 0$. Hence $\mathbb{E}[U](a_3) > \mathbb{E}[U](0)$ for all sufficiently small $a_3 > 0$, and, the equilibrium of the enlarged panel being unique, the panel is stable with respect to the new expert. □

A.5 Numerical computations for symmetric panels

I consider the symmetric panels with $x_1 = 0.4$, $x_2 = 0.6$ and $a_1 = a_2 \in \{0.3, 0.8\}$, and with $x_1 = 0.3$, $x_2 = 0.7$ and $a_1 = a_2 = 0.5$, with a uniform prior and quadratic loss. In all three, $y_0 = 1/2$ is the unique equilibrium and $T = 0$. For each panel I add one expert with an agenda on a grid of 401 points in $[0, 1]$ and a quality in $\{0.01, 0.05, 0.1, 0.2, \ldots, 0.9, 0.95, 0.99, 0.999\}$,

find all equilibria of the enlarged panel, and compare the decision-maker's utility in each of them with her initial utility. In none of the 3×5600 cases is the utility lower. For a subset of these experts, including those whose enlarged panels have several equilibria, all equilibria and payoffs were also computed exactly: with a uniform prior and quadratic loss the left-hand side of (2) is a quadratic polynomial in y between consecutive breakpoints x_i, $2x_i$ and $2x_i - 1$, so its roots can be found in closed form.

A.6 Proof of Proposition 2

Let $q(\theta) = 1 - a'\mathbf{1}\{\theta \in \Theta'(y_1)\}$. In the enlarged panel the probability of nondisclosure at the default action y_1 is $q(\theta)P_0(\theta; y_1)$, so

$$\mathbb{E}[U_1] = \int_0^1 q(\theta)P_0(\theta; y_1)u_J(y_1, \theta)dF = U(y_1) - a' \int_{\Theta'(y_1)} P_0(\theta; y_1)u_J(y_1, \theta)dF = U(y_1) + D,$$

and $D \geq 0$ because $u_J \leq 0$. The revelation set $\Theta'(y_1) = \{\theta : |\theta - x'| \leq |y_1 - x'|\} \cap [0, 1]$ is an interval with endpoint y_1 that lies on one side of y_1 and has length at most $2|x' - y_1|$. On it $u'_J(y_1, \theta) = \pm v'(|y_1 - \theta|)$ has a constant sign, so the equilibrium condition (2) of the enlarged panel,

$$V_0(y_1) - a' \int_{\Theta'(y_1)} P_0(\theta; y_1)u'_J(y_1, \theta)dF = 0,$$

gives $a' \int_{\Theta'(y_1)} P_0(\theta; y_1)v'(|y_1 - \theta|)dF = |V_0(y_1)|$. As shown in Step 4 of the proof of Theorem 2, $v(s) \leq sv'(s) \leq 2|x' - y_1|\, v'(s)$ for $0 \leq s \leq 2|x' - y_1|$; hence $D \leq 2|x' - y_1|\, |V_0(y_1)|$. Parts (i) and (ii) follow from $\mathbb{E}[U_1] - \mathbb{E}[U_0] = U(y_1) - U(y_0) + D$ and $0 \leq D \leq 2|x' - y_1|\, |V_0(y_1)|$.

With a uniform prior, quadratic loss and $P_0(\cdot\,; y_1) \equiv P$ on $\Theta'(y_1) = [y_1 - 2d, y_1]$ (or $[y_1, y_1 + 2d]$), $d = |x' - y_1|$, we have $|V_0(y_1)| = a'P \int_0^{2d} 2s\, ds = 4a'Pd^2$ and $D = a'P \int_0^{2d} s^2 ds = \frac{8}{3}a'Pd^3 = \frac{2}{3}d\, |V_0(y_1)|$, so in this case the bound in Proposition 2 overstates the direct effect by a factor of three. □

A.7 Proof of Proposition 3

When a new expert is added to the panel, the decision-maker can keep her default action unchanged. Then the revelation sets of the initial experts remain the same, and the probability of revelation of each state either remains the same or increases, so her expected utility cannot decrease. Formally, if y_c is the initial default action and P and P' are the probabilities of nondisclosure in the initial and in the enlarged panel, then $\max_y \int_0^1 P'(\theta; y) u_J(y, \theta) dF \geq \int_0^1 P'(\theta; y_c) u_J(y_c, \theta) dF \geq \int_0^1 P(\theta; y_c) u_J(y_c, \theta) dF$, since $P' \leq P$ pointwise and $u_J \leq 0$. So by committing to the status quo default action the decision-maker can guarantee herself at least her initial utility, and commitment solves the problem of the stability of the panel.