Polarizing Persuasion[*]

Axel Anderson[†] Georgetown

Nikoloz Pkhakadze[‡] Georgetown

November 29, 2021

Very Preliminary

Abstract

This paper considers a Bayesian persuasion game between a single sender and two receivers. The sender's payoff is monotone in the receivers' beliefs about the *payoff relevant state*. All agents share a common prior about this state. However, we allow disagreement about a *payoff irrelevant state*, a binary variable that enters no utility functions. When all agents have the same prior beliefs about payoff irrelevant state, then the model is a version of Kamenica and Gentzkow (2011). However disparate priors significantly change the results. In particular, it is no longer true that sender fully conceals the state when her payoff is concave in beliefs and fully reveals the state when payoff is convex in beliefs. In fact, if the sender's payoff is differentiable and strictly monotone, then even slight disagreement on the payoff irrelevant state guarantees that the sender can strictly increase her payoff by using an informative signal. Moreover, the sender's payoff is strictly increasing in the prior disagreement between the receivers.

Given extreme prior disagreement between the receivers, we show that persuasion induces significant belief polarization for two general classes of payoff functions. Specifically, if the sender's payoff is biconcave and submodular, then signals are *strongly polarizing*; namely, the signal that makes receiver 1 most confident that the payoff relevant state is 1 makes receiver 2 least confident that the payoff relevant state is 1, and the signal that makes receiver 1 second most confident the payoff relevant state is 1 makes receiver 2 second most confident the payoff relevant state is 1, etc. If the sender's payoff biconvex, then the sender chooses a message service with three signals, and receiver's beliefs are diametrically opposed for two of the three signals.

[*] We thank Luca Anderlini and Roger Lagunoff;

[†] Email: aza@georgetown.edu; web page: www.sites.google.com/view/axel-anderson

[‡] Email: np456@georgetown.edu; web page: www.sites.google.com/view/nikoloz-pkhakadze

1 Introduction

There is vast descriptive literature that documents an increase in polarization of attitudes within countries in recent decades. In addition, experiments in economics and psychology document polarization in the laboratory. To fix ideas, assume we are interested in beliefs about the extent to which human activity has contributed to climate change. We administer a survey to assess prior beliefs in a group of subjects. Subjects then get exactly the same set of scientific studies to read. We then administer the same survey and discover that beliefs have polarized. Specifically, individuals who believed that climate change was "less than 50% caused by human activity" in the first survey decrease their estimate of the impact of humans on the climate in the second survey. While individuals who believed that climate change was "more than 50% caused by human activity" in the first survey increase their estimate of the impact of humans on the climate in the second survey.

Such polarization can be explained by invoking confirmatory bias: subjects concentrate on studies that support their existing beliefs, but do not account for this selection effect when they revise their beliefs. Alternatively, we could invoke affiliation theory: subjects interact with each other and seek to align their beliefs more closely with other people with similar prior beliefs. But neither of these explanations is required to explain polarization. Two individuals could read exactly the same studies, process any new information using Bayes' rule, and end up becoming more polarized. As Benoit and Dubra (2019) show, the key to understanding such *rational polarization* is to recognize that we live in a complex multidimensional world, but measure beliefs on a smaller (typically one-dimensional) subset. For an intuition, assume the individuals in the study have different views about biases in the scientific literature. If individual A believes that climate scientists are biased toward ascribing increases in global temperatures to human activity and individual B believes that climate scientists are biased against such findings, then it can be perfectly rational for these individuals to read the same studies and have A become more convinced that global warming is not caused by human activity and B become more convinced that it is.

In Benoit and Dubra (2019), the information source (aka signal distribution) is *exogenous.* The current paper *endogenizes* the signal distribution. Specifically, we consider a Bayesian persuasion game between a single sender and a pair of receivers. As in Kamenica and Gentzkow (2011), the sender chooses an experiment (message service), but cannot directly control the public outcome of the experiment, and the receivers passively update their beliefs using Bayes' rule. The sender's payoff is a monotone function of each receiver's belief on a binary *payoff relevant* state. All agents share a common prior on this payoff relevant state. However, we allow for disparate beliefs on the *payoff irrelevant* state, a binary variable that enters no utility functions, but nonetheless may affect the receivers interpretation of the public signal.

As in the standard persuasion model, if all agents have the same prior beliefs, then the sender reveals no information when her payoff function is concave. But this is no longer true once we allow for disparate priors on the payoff irrelevant state. Toward understanding why, note that when the sender's prior differs from either (or both) receivers' prior, then the sender can design an experiment that reveals information and shifts receiver beliefs up *from the sender's perspective*. We call such upward shifts in beliefs the *misinformation effect*. The size of the information effect grows as the gulf between the receivers' prior beliefs widen, which increases the sender's expected payoff (Proposition 1). This increase is strict if the sender's payoff function is smooth and strictly increasing, as long as perfectly revealing the state is not optimal (Lemma 2). But then perfectly concealing messages can never be optimal for a sender with a smooth strictly increasing payoff function (Corollary 1).

Since receivers disagree about the payoff irrelevant state, the sender can design experiments that induce *polarization* on the payoff relevant dimension; namely experiments in which the two receivers disagree about the ordinal meaning of signals. A pair of signals j, j' is *a polarizing pair* if receiver 1's posterior is higher following signal j than j', while receiver 2's posteriors have the opposite ordering. As a preliminary step, we consider polarization in the limit case when the receivers' posteriors on the payoff irrelevant state maximally disagree. Formally, this case is equivalent to a private communication model, i.e., when the sender can design separate experiments for each receiver. When the sender has a bi-convex payoff function, Proposition 2 explicitly solves for the optimal message service. This distribution has three distinct signals. One signal will reveal that the payoff relevant state is the one that is best for the sender (state 1). The other two signals will leave one receiver sure that the state is 1 and the other sure that the state is zero. Notice that these latter two signals are a perfectly polarizing pair.

When the sender has a bi-concave and SBM payoff function, Proposition 3 establishes that optimal signals in the private communication case are *strongly polarizing*; namely, there are at least two distinct pairs of posteriors and every pair of signals is a polarizing pair. That is, the signal that makes receiver 1 most confident that the state is 1 makes receiver 2 least confident that the state is 1, and the signal that makes receiver 1 second most confident the state is 1 makes receiver 2 second most confident the state is 1, etc.

Extreme ex ante disagreement is not necessary for polarization. In Section 7, we show that polarization, including strong polarization, obtains over a wide range of receiver beliefs in numerical examples.

Related Literature. This project combines and contributes to two strands of the literature - polarization and Bayesian persuasion.

Polarization of beliefs on various issues is well documented and widely studied. For example, Lord and Ross (1979) show that attitudes toward the deterrent effect of the death penalty became more extreme after participants were exposed to the same information. Alesina, Miano, and Stantcheva (2020) documents increased polarization in subjects observing the same evidence on a number of debatable political topics. Some theoretical models capture polarization by assuming that people are ambiguity averse as in Baliga, Hanany, and Klibanoff (2013) or are biased in their processing of information, as in Fryer Jr, Harms, and Jackson (2019) and Rabin and Schrag (1999).

This is not the first paper to consider *rational* polarization. Benoit and Dubra (2019), Jern, Chang, and Kemp (2014), Loh and Phelan (2019), Andreoni and Mylovanov (2012) assume that agents optimally process new information, but allow for polarization by assuming that people have different models of how information is generated, which is modeled by assuming disparate beliefs on a non directly relevant state. All of these papers assume that the information source is exogenously specified.

Our paper extends the canonical model from Kamenica and Gentzkow (2011) in two directions - the number of receivers and heterogeneous beliefs. Ricardo and Camara (2016) retains the assumption of a single receiver, but allows the sender and receiver to have disparate prior beliefs. They show how the uncommon prior model can be transformed into a common prior problem by changing the sender's objective function. We apply their techniques to our model in Appendix A.

Pkhakadze (2021) also considers a communication game between a sender and two receivers, allowing for differing priors on a payoff irrelevant state. However, Pkhakadze (2021) analyzes cheap talk as in Crawford and Sobel (1982). Jeong (2019) considers a cheap talk model in which a politician (sender) polarizes voters (receivers). In contrast to the current work, Jeong (2019) assumes that receivers differ in their state contingent preferences, not in their prior beliefs.

Several papers explore foundations for disparate priors in economic models, including Morris (1994) and Morris (1995), and more recently Van den Steen (2010a) and Van den Steen (2010b). We do not offer micro foundations for our players disagreement. Not to contradict agreement theorem in Aumann (1976) we simply assume that players of our game agree to disagree.

Section 2 introduces the model. Section 3 presents an example in which a defense attorney attempts to persuade a jury that her client is innocent. Section 4 shows that the sender's expected payoff rises in receiver ex ante disagreement. Section 5 explores polarization when the sender's payoff function is bi-convex, while section 6 explores the bi-concave payoff case. We numerically explore the comparative statics of polarization in Section 7. Proofs immediately follow results or appear in an Appendix.

2 Model

This section describes the public communication game between one sender and two receivers.[1] The state is two dimensional $(\theta, \omega) \in \{\theta_0, \theta_1\} \times \{\omega_0, \omega_1\}$, and all player's share a common prior $p \in (0, 1)$ that $\theta = \theta_1$. The prior beliefs that $\omega = \omega_1$ are $q_s, q_1, q_2 \in (0, 1)$ for the sender, receiver one, and receiver two with $q_1 < q_s < q_2$. The sender knows the receiver's beliefs and can costlessly commit to any finite *message service*, M; namely, a finite set of signals j and state contingent probabilities $\pi^j(\theta, \omega)$. Let $\mathcal{M}$ be the space of such message services. For any message service M, the sender's subjective belief that signal j is realized thus:

$$\Pi^j_s \equiv p\left(q\pi^j(\theta_1, \omega_1) + (1-q)\pi^j(\theta_1, \omega_0)\right) + (1-p)\left(q\pi^j(\theta_0, \omega_1) + (1-q)\pi^j(\theta_0, \omega_0)\right) \quad (1)$$

and receiver i's subjective belief is analogously defined given q_i as Π^j_i.

One signal is drawn from the chosen message service, and this signal is commonly observed by the sender and each receiver. Let P^j_s (P^j_i) be the posterior belief of the sender (receiver i) after observing signal j. We assume that all agents update their beliefs according to Bayes' rule using their own prior beliefs; and thus:

$$P^j_k = \frac{p\left(q_k\pi^j(\theta_1, \omega_1) + (1-q_k)\pi^j(\theta_1, \omega_0)\right)}{\Pi^j_k} \qquad \forall\ k \in \{s, 1, 2\} \quad (2)$$

For now we abstract from receiver choices by assuming that the sender's C^2 payoff V only depends on the receivers' posterior beliefs (P^j_1, P^j_2). The sender's maximization problem is thus:

$$V^*(p, q_s, q_1, q_2) = \max_{M \in \mathcal{M}} \sum_j \Pi^j_s V(P^j_1, P^j_2) \qquad \text{s.t. (1) and (2)} \quad (3)$$

While the sender's payoff function does not directly depend on receiver beliefs in state ω, the indirect payoff function V^* does. This owes to the fact that the posterior beliefs P^j_i on θ depend on the prior belief q_i on ω, and the fact that the sender evaluates the probabilities Π^j_s with her own prior q_s. Altogether, while it would be more precise to refer to ω as *not directly* payoff relevant, we henceforth refer to ω as the *payoff*

[1] The model embeds the information structure in Benoit and Dubra (2019) in a Kamenica and Gentzkow (2011) communication game. Critically for our purposes, Ricardo and Camara (2016) allow for heterogeneous prior beliefs.

irrelevant state and θ as the *payoff relevant state*. We assume that V is non-decreasing in each argument, which trivially implies that V^* is non-decreasing in p.

3 Example: A Zealous Defender

In this section we consider a defense attorney trying to persuade a jury consisting of two receivers. The defendant is either guilty (θ_0) or innocent (θ_1). The payoff irrelevant state ω can represent how the jurors interpret evidence presented at trial. For example, it could parameterize juror beliefs about the reliability of eyewitness testimony. In order to streamline the analysis, we assume $q_2 = 1 - q_1 > 1/2$.

The defense attorney's trial strategy determines the message service, and the trial itself is the signal generated by the trial strategy. After observing the trial (signal), each juror votes for conviction if her posterior belief on guilt P_i exceeds 1/2, and votes to acquit otherwise. A conviction requires unanimity. For simplicity, we assume that the defense attorney gets a payoff of 1 from acquittal and 0 from conviction.

For sake of comparison with the standard case, restrict attention to message services with signal distributions that are independent of the payoff irrelevant state. This is equivalent to the standard model with a single receiver, since this restriction implies that the jurors will have the same posterior belief $P_1^j = P_2^j$ following any signal j. In this case, the defense attorney chooses to reveal no information when $p \geq 1/2$. When $p < 1/2$, the defense attorney chooses a message service that always sends an "innocent" signal when the defendant is innocent, and also sends an "innocent" signal with chance $p/(1-p)$ when the defendant is guilty. That is, a binary signal message service with state contingent probabilities:

$$\begin{array}{ccc} \pi^1 : & & \\ & \theta_0 & \theta_1 \\ \omega_0 & \frac{p}{1-p} & 1 \\ \omega_1 & \frac{p}{1-p} & 1 \end{array} \qquad\qquad \begin{array}{ccc} \pi^2 : & & \\ & \theta_0 & \theta_1 \\ \omega_0 & \frac{1-2p}{1-p} & 0 \\ \omega_1 & \frac{1-2p}{1-p} & 0 \end{array} \tag{4}$$

Message service (4) induces posteriors $P_1^1 = P_2^1 = 1/2$ and $P_1^2 = P_2^2 = 0$ (Figure 1, left). The defense attorney chooses the misreporting chance $p/(1-p)$ so that the good signal barely convinces the jurors to acquit and leaves them sure the defendant is guilty following the "guilty" signal. Thus, the acquittal chance is $p(1)+(1-p)\cdot p/(1-p) = 2p$. Altogether, the ex ante expected value for the sender is $V^* = \min\{1, 2p\}$.

When $p < 1/2$ the sender can strictly improve her payoff by conditioning on the payoff irrelevant state. Specifically, define $\lambda(p, q_2) \equiv \min\{1, pq_2[(1-p)(1-q_2)]^{-1}\}$,

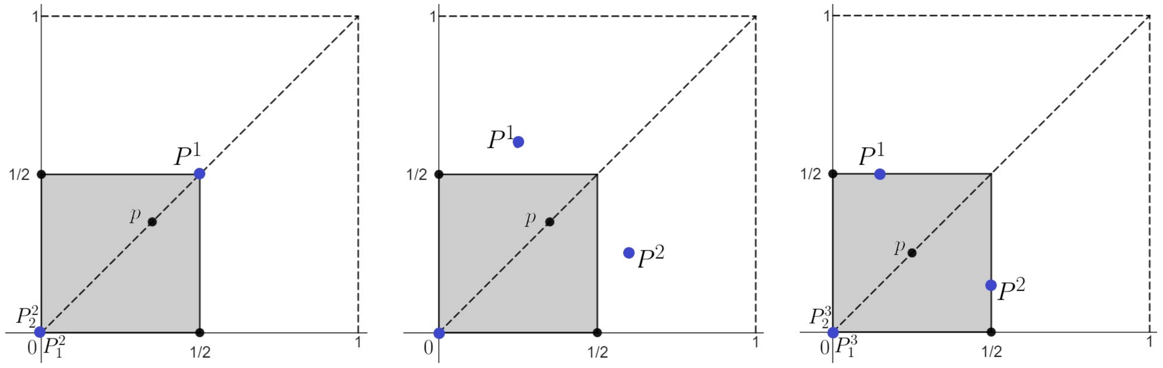

Figure 1: **Posterior Beliefs for a Zealous Defense.** In all graphs the sender has value $V = 0$ on the grey shaded region and $V = 1$ elsewhere. The left graph depicts the optimal posterior beliefs given prior $p < 1/2$, when the sender is prevented from choosing a message service that conditions on the payoff irrelevant dimension. When $p \in [1-q_2, 1/2]$, the sender can guarantee $V = 1$ by conditioning on the payoff irrelevant dimension to induce posterior beliefs as shown in the middle graph. For lower values of p, the optimal message service has three signals, as illustrated on the right.

then the following message service is *optimal*:

$$\pi^1: \begin{array}{ccc} & \theta_0 & \theta_1 \\ \omega_0 & \lambda(p,q_2) & 0 \\ \omega_1 & 0 & 1 \end{array} \qquad \pi^2: \begin{array}{ccc} & \theta_0 & \theta_1 \\ \omega_0 & 0 & 1 \\ \omega_1 & \lambda(p,q_2) & 0 \end{array} \qquad \pi^3: \begin{array}{ccc} & \theta_0 & \theta_1 \\ \omega_0 & 1-\lambda(p,q_2) & 0 \\ \omega_1 & 1-\lambda(p,q_2) & 0 \end{array} \tag{5}$$

To gain some insight into this new message service, first assume $p \geq 1 - q_2$, and thus, $\lambda(p, q_2) = 1$. In this case, the third signal is never sent and by $q_2 = 1 - q_1$ we have:

$$P_2^1 = P_1^2 = \frac{pq_2}{pq_2 + (1-p)(1-q_2)} \quad \text{and} \quad P_2^2 = P_1^1 = \frac{p(1-q_2)}{p(1-q_2) + (1-p)q_2}$$

Straightforward algebra establishes that $p \geq 1 - q_2$ implies that $P_2^1 = P_1^2 \geq 1/2$; and thus, the defendant is guaranteed to get at least one vote for acquittal (see Figure 1, middle). Recalling that $1 - q_2 < 1/2$, this means that the defense attorney can guaranteed acquittal for a wider range of prior beliefs than with the binary message service (4). Notice that message service (5) results in *polarization in juror beliefs*. Specifically, if signal 1 is sent, then juror 1 becomes more convinced that the defendant is guilty, while juror 2 become more convinced that the defendant is innocent. If signal 2 is sent, the opposite occurs: juror 1 interprets this as good news for the defendant and juror 2 interprets it as bad news for the defendant. In other words, this message service induces a strong form of *polarization*: all signals cause the jurors' beliefs to move in opposite directions on the payoff relevant dimension.

If $p < 1 - q_2$, then $\lambda(p, q_2) < 1$ and $P_2^1 = P_1^2 = 1/2 > p$ and $P_1^1 = P_2^2 < p$. Thus, following signals 1 and 2: (a) juror beliefs polarize and (b) one juror votes for acquittal. But if signal 3 is sent, then both jurors are sure the defendant is guilty and vote to convict (see Figure 1, right). Thus, the chance of acquittal is equal to the chance that signal 3 is not sent: $1 - (1 - p)(1 - \lambda(p, q_2)) = p/(1 - q_2)$. This is again strictly higher than $2p$, the chance of acquittal with message service (4). In summary: the defense attorney increases her expected payoff by exploiting juror ex ante disagreement on the payoff irrelevant state, inducing polarized posterior beliefs on the payoff relevant state.

The next section establishes that the sender always benefits from ex ante disagreement, and that this benefit is monotone in the size of the disagreement.

4 The Sender Benefits from Ex Ante Disagreement

The constraint set in sender maximization (3) only enters the objective function indirectly. The standard approach in the Persuasion literature is to allow the sender to choose the distribution over receiver posteriors subject to a constraint that takes account of receivers' posterior beliefs and Bayesian updating. Toward this end, define $\alpha \in \mathbb{R}$ as the unique solution to $q_s = \alpha q_1 + (1 - \alpha) q_2$ (valid by $q_1 \neq q_2$ and all beliefs bounded away from 0 and 1). Then the following Lemma fruitfully reformulates the sender's maximization problem.

Lemma 1. *A message service solves the sender's maximization* (3)*, if and only if, it induces distributions over beliefs that solves:*

$$V^* = \max_{\Pi_i, P_i} \sum_j \left(\alpha \Pi_1^j + (1 - \alpha) \Pi_2^j\right) V(P_1^j, P_2^j) \quad \textit{s.t.:} \tag{6}$$

$$\sum_j \Pi_i^j P_i^j = p \quad \forall i \tag{7}$$

$$\frac{q_1}{q_2} \leq \frac{P_1^j \Pi_1^j}{P_2^j \Pi_2^j} \leq \frac{1 - q_1}{1 - q_2} \quad \textit{and} \quad \frac{q_1}{q_2} \leq \frac{(1 - P_1^j)\Pi_1^j}{(1 - P_2^j)\Pi_2^j} \leq \frac{1 - q_1}{1 - q_2} \quad \forall P^j \notin \{\boldsymbol{0}, \boldsymbol{1}\} \tag{8}$$

If we set $q_s = q_1$, and then take the limit $q_1 \to q_2$, the reformulation in Lemma 1 is a special case of the reformulation in Kamenica and Gentzkow (2011).[2] In particular, when $q_1 = q_2$, constraint (8) demands that the distribution over receiver beliefs must be

[2] The maximization in Lemma 1 is not well defined in the limit $q_1 = q_2 \neq q$, since $\alpha = (q_2 - q)/(q_2 - q_1)$. When $q_1 = q_2 \neq q$ the original maximization (3) is a special case of Ricardo and Camara (2016). They provide an alternative reformulation for this case, which we apply to our model in Section A of Appendix. This result also allows us to restrict attention to message service with at most 4 signals

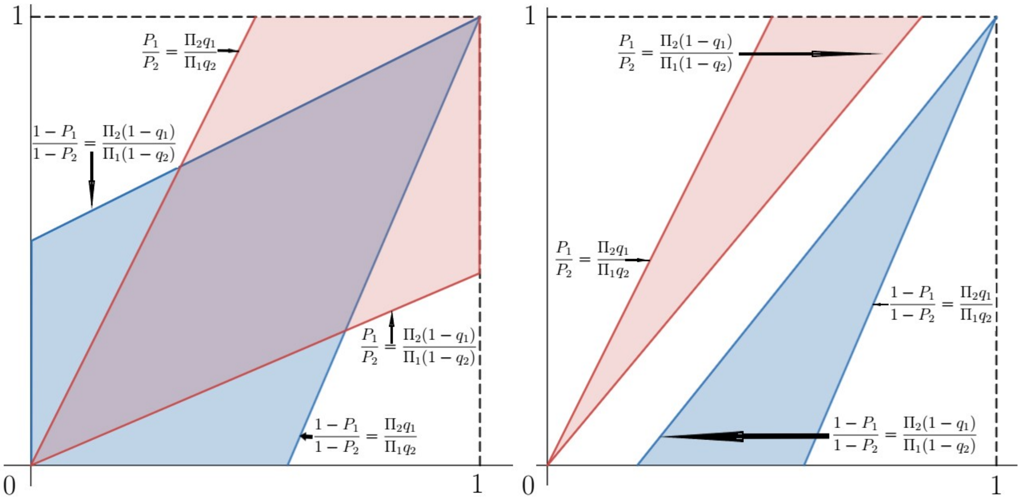

Figure 2: **Feasible Belief Pairs.** The joint restrictions on beliefs (P_1^j, P_2^j) implied by (8) is the shaded region in the left graph. On the right, we assume that both the upper and lower bound of left hand constraints in (8) are above the 45 degree line, which then implies that the upper and lower bound of the right hand constraints in (8) are below the 45 degree line. But then pairs must be in both shaded regions in the figure, i.e. the intersection of the four inequalities is empty, a contradiction.

identical. Further, $q_s = q_1 = q_2$ implies $\alpha = 1$, so the sender's expected payoff is evaluated with this common belief distribution. Altogether, in this limit the sender chooses a common belief distribution for all players, subject to the martingale property (7).

Constraint (8) is a consistency requirement on the joint distribution of receiver beliefs and is novel to the current model, constraining the divergence in receiver posterior beliefs on the payoff relevant state θ. The left graph in Figure 2 illustrates the convex set of allowable posterior pairs (P_1^j, P_j^2) implied by (8) for fixed (Π_1^j, Π_2^j) for any feasible message service. In particular, for any feasible message service, the constraints in (8) must "straddle the 45 degree line;" namely the upper and lower bounds of the constraints obey:

$$\begin{array}{llll} \frac{P_1^j \Pi_1^j}{P_2^j \Pi_2^j} = \frac{q_1}{q_2} \quad \Rightarrow \quad \frac{P_1^j}{P_2^j} \leq 1 & \text{and} & \frac{P_1^j \Pi_1^j}{P_2^j \Pi_2^j} = \frac{1-q_1}{1-q_2} \quad \Rightarrow \quad \frac{P_1^j}{P_2^j} \geq 1 \\ \frac{(1-P_1^j)\Pi_1^j}{(1-P_2^j)\Pi_2^j} = \frac{q_1}{q_2} \quad \Rightarrow \quad \frac{P_1^j}{P_2^j} \geq 1 & \text{and} & \frac{(1-P_1^j)\Pi_1^j}{(1-P_2^j)\Pi_2^j} = \frac{1-q_1}{1-q_2} \quad \Rightarrow \quad \frac{P_1^j}{P_2^j} \leq 1 \end{array}$$

Recalling that $q_1 < q_2$, *receiver ex ante disagreement increases* when q_1 decreases and/or q_2 rises. Since either change relaxes constraint (8), Lemma 1 implies:

Proposition 1. *The sender's expected value V^* is non-decreasing in receiver ex ante disagreement.*

When is V^* strictly increasing in ex ante disagreement? Clearly, a robust necessary condition is that V is strictly increasing. It turns out that this is also sufficient if V is continuously differentiable and optimal message services are not *perfectly revealing*, i.e. inducing posterior beliefs $((0,0),(1,1))$ with chances $(1-p,p)$.

Lemma 2. *Assume V is strictly increasing and continuously differentiable. At least one of the constraints in* (8) *holds with equality for any signal j that does not reveal the state, i.e., with $(P_1^j,P_2^j)\in(0,1)^2$. Consequently, V^* is strictly increasing in ex ante disagreement when perfectly revealing message services are not optimal.*

Step 1: If $(P_1^j,P_2^j)\in(0,1)^2$ and (8) does not bind, then $E_s[V(P_1,P_2)]<V^*$.

Proof: Toward a contradiction, assume an optimal distribution over posteriors such that all constraints in (8) are slack and $(P_1^j,P_2^j,\Pi_1^j,\Pi_2^j)$ with $P^j\notin\{\mathbf{0},\mathbf{1}\}$. Now consider a new distribution over posteriors that is identical, except posteriors (P_1^j,P_2^j) are replaced by three new pairs $(P_1^j+\varepsilon,P_2^j+\varepsilon)$, $(P_1^j+\varepsilon,P_2^j-\varepsilon)$, and $(P_1^j-\varepsilon,P_2^j+\varepsilon)$ with associated receiver probabilities $(\delta\Pi_1^j,\delta\Pi_2^j)$, $((1/2-\delta)\Pi_1^j,\Pi_2^j/2)$, and $(\Pi_1^j/2,(1/2-\delta)\Pi_2^j)$. By construction, if the original distribution satisfied Bayesian constraint (7), then so does this new distribution. And for sufficiently small (ε,δ), all constraints in (8) will be met for all three of these new pairs of beliefs and receiver probabilities.

Now take a first order approximation to the sender's value at each pair of beliefs:

$$\begin{aligned}
V(P_1^j+\varepsilon,P_2^j+\varepsilon) &\approx V(P_1^j,P_2^j)+\varepsilon\left(V_1(P_1^j,P_2^j)+V_2(P_1^j,P_2^j)\right)\\
V(P_1^j+\varepsilon,P_2^j-\varepsilon) &\approx V(P_1^j,P_2^j)+\varepsilon\left(V_1(P_1^j,P_2^j)-V_2(P_1^j,P_2^j)\right)\\
V(P_1^j-\varepsilon,P_2^j+\varepsilon) &\approx V(P_1^j,P_2^j)+\varepsilon\left(V_2(P_1^j,P_2^j)-V_1(P_1^j,P_2^j)\right)
\end{aligned}$$

And use these approximations to find the first order change in sender value $E_s[V(P_1,P_2)]$ given this change in the belief distribution:

$$\begin{aligned}
\Delta E_s[V] &\approx \varepsilon\left[\delta\left(\alpha\Pi_1^j+(1-\alpha)\Pi_2^j\right)(V_1+V_2)+\alpha\delta\Pi_1^j(V_2-V_1)+(1-\alpha)\delta\Pi_2^j(V_1-V_2)\right]\\
&= 2\varepsilon\delta\left(\alpha\Pi_1^jV_2+(1-\alpha)\Pi_2^jV_1\right)>0 \qquad (9)
\end{aligned}$$

Since this is strictly positive, the first distribution could not have been optimal.

Step 2: V^* strictly rises in ex ante disagreement when perfect revelation is not optimal.

Proof: If perfect revelation is not optimal for prior beliefs (q_1,q_2), then any optimal distribution $(\Pi_1^*,\Pi_2^*,P_1^*,P_2^*)$ contains a signal j, such that $P^j\notin\{\mathbf{0},\mathbf{1}\}$. Now consider the sender's maximization with prior beliefs (q_1',q_2') obeying $q_1'\le q_1$ and $q_2'\ge q_2$ with at least one strict inequality. Notice that $(\Pi_1^*,\Pi_2^*,P_1^*,P_2^*)$ remains feasible at (q_1',q_2'), and none of the constraints in (8) bind at signal j. Thus, by Step 1, we must have $V^*(q_1', q_2') > V^*(q_1, q_2)$. □

Intuitively, if the consistency constraints do not bind, then the sender can take advantage of the receiver's disparate beliefs on the payoff irrelevant state to increase $\mathrm{E}_S[P_i]$ for $i \in \{1, 2\}$, i.e. the sender's expectation of the each receiver's posterior. For sufficiently small changes, this first order *misinformation effect* dominates any second order effects.

A message service is *perfectly concealing* if both receivers posterior beliefs are equal to their prior beliefs with probability 1. Notice that all constraints in (8) are slack for any perfectly concealing message service (by $p \in (0, 1)$). Thus, Lemma 2 yields the following immediate corollary.

Corollary 1. *Perfectly concealing message services are never optimal when V is strictly increasing and continuously differentiable.*

Simply put, in order to benefit from the receiver disagreement on the payoff irrelevant state, the sender must use a message service that contains some information. At first glance this seems to be in conflict with the standard result that perfectly concealing messages are optimal for concave V. To reconcile the results, assume $q_s = q_2$ and then consider the limit $q_1 \to q_2$. Notice in this limit that constraint (8) implies that $P_1^j = P_2^j$, and polarization must vanish in this limit. In this limit, the sender will then choose to conceal when $V(P, P)$ is concave in P. In other words, the standard result emerges in the common prior limit of the current model.

Now consider the sender's incentives to reveal the state. A message service is *partially revealing* if it contains a signal j sent with positive probability such that $P^j \in \{(0, 0), (1, 1)\}$. Toward sufficient conditions for partial revelation, we say that the sender's value is *bi-convex* when $V(P_1, P_2)$ convex in P_1 for all P_2 and convex in P_2 for all P_1, *(strictly) supermodular* (SPM) when

$$V(P_1, \hat{P}_2) + V(P_1, \hat{P}_2) < (\leq) V(P_1, P_2) + V(\hat{P}_1, \hat{P}_2) \quad \forall\ (\hat{P}_1, \hat{P}_2) > (P_1, P_2)$$

and (strictly) submodular when $-V$ is (strictly) supermodular. Then we have, the following partial revelation result.

Conjecture 1. *Optimal message services are partially revealing if V is bi-convex and SPM.*

In the standard model, convexity of the sender's payoff function V is sufficient for message services to be *perfectly revealing*, i.e., reveal the state with probability 1. Since full revelation is a special case of partial revelation, Conjecture 1 does not contradict full revelation being optimal for convex V. However, we show in the next section that V convex is not sufficient for full revelation. In fact, given sufficient ex-ante disagreement, strict convexity of V *guarantees that full revelation is* ***not*** *optimal.*

5 Polarization for Bi-Convex Values

We now turn to our focus: understanding when the sender induces polarization. We will show that the sender induces substantial polarization when she has a bi-convex value function, as long as the receivers posterior beliefs on the payoff irrelevant state are sufficiently far apart.

Consider the limit $q_1 \to 0$ and $q_2 \to 1$ (and thus, $\alpha \to 1 - q_s$). By Proposition 1, the sender's value is maximized in this limit. Notice that the consistency constraint (8) vanishes in this limit; and thus, the sender's choice of posterior distributions for one receiver is unconstrained by her choice of posterior distributions for the other receiver. In other words, this limit case is formally equivalent to separate *private communication* with each receiver on the payoff relevant state.[3] In a persuasion model with one receiver, the sender chooses to fully reveal the state when her value is convex in receiver beliefs. Here bi-convexity implies a specific trinomial distribution over extremal posterior beliefs in the private communication model.

Proposition 2. *Assume private communication. If V is strictly bi-convex and increasing, then the optimal message service must induce posterior beliefs $(P_1, P_2) = ((1,0),(0,1),(1,1))$ with probabilities $(\Pi_1, \Pi_2) = ((0, 1-p),(1-p,0),(p,p))$.*

Intuitively, bi-convexity of beliefs induces the sender to restrict attention to belief distributions supported on $\{(0,0),(1,0),(0,1),(1,1)\}$. The given distribution first order dominates all other distributions on this support from the point of view of the sender subject to the constraint that $E_i[P_i] = p$ (i.e. each receiver does not expect his belief to change).

This belief distribution is not completely polarizing, since it places positive weight on the perfectly correlated beliefs $(P_1, P_2) = (1,1)$. However, two of the three signals yield beliefs that are maximally polarized. More specifically, from the sender's ex ante point of view, there is a $1-p$ chance that posterior beliefs end up being maximally polarized. A cardinal measure of polarization is the covariance between P_1 and P_2, with more negative covariance corresponding to more polarization. Calculating the covariance (from the sender's point of view) for the distribution in Proposition 2 we

[3] More generally, we conjecture that our general model is formally equivalent to a model in which the sender attempts to privately communicate, but there's some chance that signals are publicly revealed. Formal analysis in progress.

get $-q_s(1-q_s)(1-p)^2 < 0$. Thus, this measure of polarization is maximized at $q_s = 1/2$ and decreasing in p.[4]

To show how the sender can induce this distribution over beliefs, consider a three signals message service with state contingent probabilities:

$$\begin{array}{ccc} \pi^1: & & \\ & \theta_0 & \theta_1 \\ \omega_0 & 0 & \varepsilon \\ \omega_1 & 1 & 0 \end{array} \qquad \begin{array}{ccc} \pi^2: & & \\ & \theta_0 & \theta_1 \\ \omega_0 & 1 & 0 \\ \omega_1 & 0 & \varepsilon \end{array} \qquad \begin{array}{ccc} \pi^3: & & \\ & \theta_0 & \theta_1 \\ \omega_0 & 0 & 1-\varepsilon \\ \omega_1 & 0 & 1-\varepsilon \end{array} \tag{10}$$

Fixing $q_1 = 0$ and $q_2 = 1$ and taking the limit $\varepsilon \to 0$, yields the desired belief distribution.[5]

6 Polarization with Bi-Concave Values

In this section relate the second order properties of the senders value V to payoff improving transformation of the distributions over beliefs. We then give sufficient conditions for a very strong form of polarization for the case of sufficient ex ante disagreement among the receivers.[6]

6.1 Mean Improving Contractions

Let $P_i^L < P_i^H$ for $i \in \{1,2\}$ with $(\Delta_i^H, \Delta_i^L) \geq 0$ (strictly positive for at least one i) and $P_i^H - P_i^L \geq \Delta_i^L + \Delta_i^H$. A *mean improving positive contraction* moves probability mass $\varepsilon > 0$ in the sender's subjective distribution over receiver beliefs from posterior pair (P_1^L, P_2^L) to posterior pair $(P_1^L + \Delta_1^L, P_2^L + \Delta_2^L)$ and mass $\delta > 0$ from posterior pair (P_1^H, P_2^H) to posterior pair $(P_1^H - \Delta_1^H, P_2^H - \Delta_2^H)$, such that the sender's posterior beliefs weakly increase, i.e. $\varepsilon\Delta_i^L \geq \delta\Delta_i^H$ for $i \in \{1,2\}$. A *mean improving negative contraction* moves probability mass $\varepsilon > 0$ in the sender's subjective distribution from posterior pair (P_1^L, P_2^H) to posterior pair $(P_1^L + \Delta_1^L, P_2^H - \Delta_2^H)$ and mass $\delta > 0$ from posterior pair (P_1^H, P_2^L) to posterior pair $(P_1^H - \Delta_1^H, P_2^L + \Delta_2^L)$, such that the sender expects weakly higher posterior beliefs, i.e. $\varepsilon\Delta_1^L \geq \delta\Delta_1^H$ and $\delta\Delta_2^L \leq \varepsilon\Delta_2^H$. The two types of mean improving contractions are illustrated in Figure 3 (top).

[4] The same comparative statics hold if we measure polarization with the correlation coefficient between P_1 and P_2.

[5] To be added: a continuity result: V sufficiently smooth (C^2?) and strictly bi-convex implies that the optimal solution is close to the distribution in Proposition 2 (in the Lévy-Prokhorov metric) for sufficiently spread priors q_1, q_2.

[6] Characterizing polarization given any $q_1 < q_2$ is ongoing.

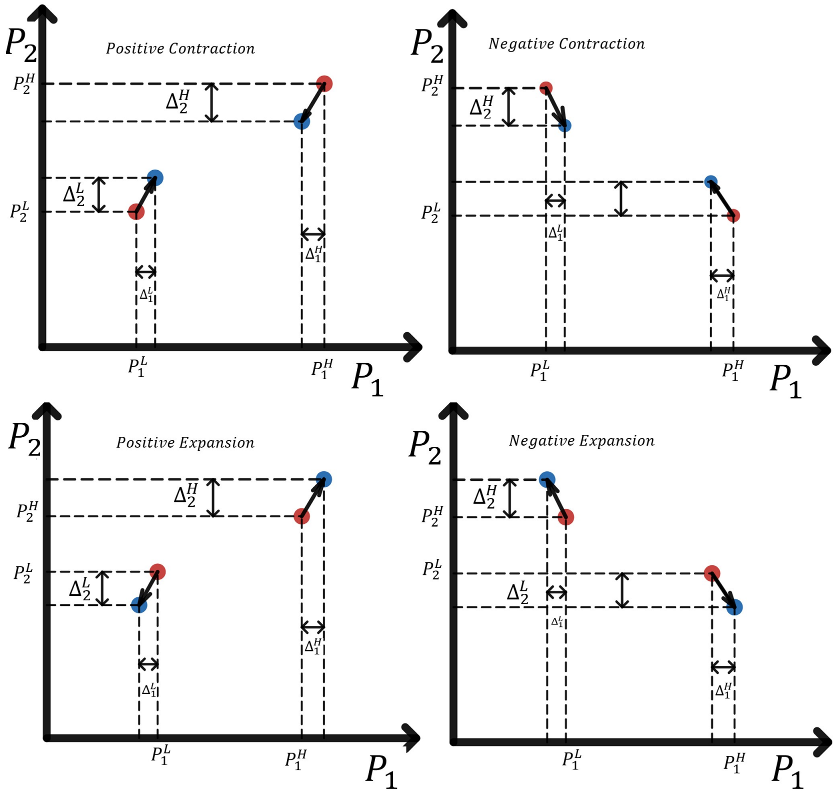

Figure 3: **Posterior Belief Swaps.** The top two graphs illustrate the two types of posterior belief contractions. The bottom two graphs illustrate the two types of posterior belief expansions

The senders value is *bi-concave* when $-V$ is bi-convex.

Lemma 3. *Assume V is bi-concave. Mean improving positive contractions increase the sender's expected payoff when V is SBM,*[7] *while mean improving negative contractions increase the sender's expected payoff when V is SPM.*

Proof: Define $P_i^{L+} = P_i^L + \Delta_i^L$ and $P_i^{H-} = P_i^H - \Delta_i^H$.

Step 1: Positive Contraction. Using in order: V bi-concave (with $P_1^{H-} \geq P_1^{L+}$), then V SBM (with $P_2^H > P_2^L$), and finally $\varepsilon\Delta_1^L \geq \delta\Delta_1^H$ (with V non-decreasing) we

[7] Muller and Scarsini (2012) captures this as a special case when the expansion is mean preserving without assuming differentiability of V.

discover:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L+},P_2^L)-V(P_1^L,P_2^L)\right]+\delta\left[V(P_1^{H-},P_2^H)-V(P_1^H,P_2^H)\right] \\
= \; & \varepsilon\int_{P_1^L}^{P_1^{L+}} V_1(s,P_2^L)ds-\delta\int_{P_1^{H-}}^{P_1^H} V_1(s,P_2^H)ds \\
\geq \; & \varepsilon\Delta_1^L V_1(P_1^{H-},P_2^L)-\delta\Delta_1^H V_1(P_1^{H-},P_2^H)\geq\left[\varepsilon\Delta_1^L-\delta\Delta_1^H\right]V_1(P_1^{H-},P_2^H)\geq 0
\end{aligned} \tag{11}$$

where the first inequality is strict when $\Delta_1^H, \Delta_1^L > 0$.

Similarly, using in order: V bi-concave (with $P_2^{H-} \geq P_2^{L+}$), then V SBM (with $P_1^{H-} \geq P_1^{L+}$), and finally $\varepsilon\Delta_2^L \geq \delta\Delta_2^H$ (with V non-decreasing) we discover:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L+},P_2^{L+})-V(P_1^{L+},P_2^L)\right]+\delta\left[V(P_1^{H-},P_2^{H-})-V(P_1^{H-},P_2^H)\right] \\
= \; & \varepsilon\int_{P_2^L}^{P_2^{L+}} V_2(P_1^{L+},t)dt-\delta\int_{P_2^{H-}}^{P_2^H} V_2(P_1^{H-},t)dt \\
\geq \; & \varepsilon\Delta_2^L V_2(P_1^{L+},P_2^{H-})-\delta\Delta_2^H V_2(P_1^{H-},P_2^{H-})\geq\left[\varepsilon\Delta_2^L-\delta\Delta_2^H\right]V_2(P_1^{H-},P_2^{H-})\geq 0
\end{aligned} \tag{12}$$

where the first inequality is strict when $\Delta_2^H, \Delta_2^L > 0$.

Now use (11) and (12) to sign the change in the sender's expected payoff from any mean improving inframodular swap:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L+},P_2^{L+})-V(P_1^L,P_2^L)\right]+\delta\left[V(P_1^{H-},P_2^{H-})-V(P_1^H,P_2^H)\right] \\
= \; & \varepsilon\left[V(P_1^{L+},P_2^L)-V(P_1^L,P_2^L)\right]+\delta\left[V(P_1^{H-},P_2^H)-V(P_1^H,P_2^H)\right] \\
& +\varepsilon\left[V(P_1^{L+},P_2^{L+})-V(P_1^{L+},P_2^L)\right]+\delta\left[V(P_1^{H-},P_2^{H-})-V(P_1^{H-},P_2^H)\right]\geq 0
\end{aligned}$$

where the inequality is strict when either strict when $\Delta_1^H, \Delta_1^L > 0$ or $\Delta_2^H, \Delta_2^L > 0$.

Step 2: Negative Contraction. Using in order: V bi-concave (with $P_1^{H-} \geq P_1^{L+}$), then V SPM (with $P_2^H > P_2^L$), and finally $\varepsilon\Delta_1^L \geq \delta\Delta_1^H$ (with V non-decreasing) we discover:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L+},P_2^H)-V(P_1^L,P_2^H)\right]+\delta\left[V(P_1^{H-},P_2^L)-V(P_1^H,P_2^L)\right] \\
= \; & \varepsilon\int_{P_1^L}^{P_1^{L+}} V_1(s,P_2^H)ds-\delta\int_{P_1^{H-}}^{P_1^H} V_1(s,P_2^L)ds \\
\geq \; & \varepsilon\Delta_1^L V_1(P_1^{H-},P_2^H)-\delta\Delta_1^H V_2(P_1^{H-},P_2^L)\geq\left[\varepsilon\Delta_1^L-\delta\Delta_1^H\right]V_1(P_1^{H-},P_2^L)\geq 0
\end{aligned} \tag{13}$$

where the first inequality is strict when $\Delta_1^H, \Delta_1^L > 0$.

Similarly, using in order: V bi-concave (with $P_2^{H-} \geq P_2^{L+}$), then V SPM (with

$P_1^{H-} \geq P_1^{L+}$), and finally $\delta\Delta_2^L \geq \varepsilon\Delta_2^H$ (with V non-decreasing) we discover:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L+}, P_2^{H-}) - V(P_1^{L+}, P_2^{H})\right] + \delta\left[V(P_1^{H-}, P_2^{L+}) - V(P_1^{H-}, P_2^{L})\right] \\
= \; & \delta \int_{P_2^L}^{P_2^{L+}} V_2(P_1^{H-}, t)dt - \varepsilon \int_{P_2^{H-}}^{P_2^{H}} V_2(P_1^{L+}, t)dt \\
\geq \; & \delta\Delta_2^L V_2(P_1^{H-}, P_2^{H-}) - \varepsilon\Delta_2^H V_2(P_1^{L+}, P_2^{H-}) \geq \left[\delta\Delta_2^L - \varepsilon\Delta_2^H\right] V_2(P_1^{L+}, P_2^{H-}) \geq 0 \quad (14)
\end{aligned}$$

where the first inequality is strict when $\Delta_2^H, \Delta_2^L > 0$.

Now use (13) and (14) to sign the change in the sender's expected payoff from any mean improving contractive swap:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L+}, P_2^{H-}) - V(P_1^{L}, P_2^{H})\right] + \delta\left[V(P_1^{H-}, P_2^{L+}) - V(P_1^{H}, P_2^{L})\right] \\
= \; & \varepsilon\left[V(P_1^{L+}, P_2^{H}) - V(P_1^{L}, P_2^{H})\right] + \delta\left[V(P_1^{H-}, P_2^{L}) - V(P_1^{H}, P_2^{L})\right] \\
& + \varepsilon\left[V(P_1^{L+}, P_2^{H-}) - V(P_1^{L+}, P_2^{H})\right] + \delta\left[V(P_1^{H-}, P_2^{L+}) - V(P_1^{H-}, P_2^{L})\right] \geq 0
\end{aligned}$$

where the inequality is strict when either strict when $\Delta_1^H, \Delta_1^L > 0$ or $\Delta_2^H, \Delta_2^L > 0$. □

6.2 Payoff Improving Expansions

Let $P_i^L < P_i^H$ for $i \in \{1, 2\}$ with $(\Delta_i^H, \Delta_i^L) \geq 0$ (strictly positive for at least one i). A *mean improving positive expansion* moves probability mass $\varepsilon > 0$ in the sender's subjective distribution over receiver beliefs from posterior pair (P_1^L, P_2^L) to posterior pair $(P_1^L - \Delta_1^L, P_2^L - \Delta_2^L)$ and mass $\delta > 0$ from posterior pair (P_1^H, P_2^H) to posterior pair $(P_1^H + \Delta_1^H, P_2^H + \Delta_2^H)$, such that the sender expects weakly higher posterior beliefs, i.e. $\delta\Delta_i^H \geq \varepsilon\Delta_i^L$ for $i \in \{1, 2\}$. A *mean improving negative expansion* moves probability mass $\varepsilon > 0$ in the sender's subjective distribution over receiver beliefs from posterior pair (P_1^L, P_2^H) to posterior pair $(P_1^L - \Delta_1^L, P_2^H + \Delta_2^H)$ and mass $\delta > 0$ from posterior pair (P_1^H, P_2^L) to posterior pair $(P_1^H + \Delta_1^H, P_2^L - \Delta_2^L)$, such that the sender expects weakly higher posterior beliefs, i.e. $\delta\Delta_1^L \geq \varepsilon\Delta_1^L$ and $\varepsilon\Delta_2^H \geq \delta\Delta_2^L$. The two types of mean improving contractions are illustrated in Figure 3 (bottom).

Lemma 4. *Assume V is bi-convex. Mean improving negative expansions increase the sender's expected payoff when V is SBM, while mean improving positive expansions increase the sender's expected payoff when V is SPM.*[8]

Proof: Define $P_i^{L-} = P_i^L - \Delta_i^L$ and $P_i^{H+} = P_i^H + \Delta_i^H$.

[8] A mean preserving positive expansion is a special case of an ultramodular transfer. Muller and Scarsini (2012) show that ultramodular transfers increase expected payoffs with any SPM and bi-convex utility function (aka any ultramodular utility function).

Step 1: Negative Expansions. Using in order: V bi-convex (with $P_1^H > P_1^L$), then V SBM (with $P_2^H > P_2^L$), and finally $\delta\Delta_1^H \geq \varepsilon\Delta_1^L$ (with V non-decreasing) we discover:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L-}, P_2^H) - V(P_1^L, P_2^H)\right] + \delta\left[V(P_1^{H+}, P_2^L) - V(P_1^H, P_2^L)\right] \\
= \; & \delta \int_{P_1^H}^{P_1^{H+}} V_1(s, P_2^L)ds - \varepsilon \int_{P_1^{L-}}^{P_1^L} V_1(s, P_2^H)ds \\
\geq \; & \delta\Delta_1^H V_1(P_1^L, P_2^L) - \varepsilon\Delta_1^L V_1(P_1^L, P_2^H) \geq \left[\delta\Delta_1^H - \varepsilon\Delta_1^L\right] V_1(P_1^L, P_2^H) \geq 0 \qquad (15)
\end{aligned}$$

where the first inequality is strict when $\Delta_1^H, \Delta_1^L > 0$.

Similarly, using in order: V bi-convex (with $P_2^H > P_2^L$), then V SBM (with $P_1^{H+} > P_1^{L-}$), and finally $\varepsilon\Delta_2^H \geq \delta\Delta_2^L$ (with V non-decreasing) we discover:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L-}, P_2^{H+}) - V(P_1^{L-}, P_2^H)\right] + \delta\left[V(P_1^{H+}, P_2^{L-}) - V(P_1^{H+}, P_2^L)\right] \\
= \; & \varepsilon \int_{P_2^H}^{P_2^{H+}} V_2(P_1^{L-}, t)dt - \delta \int_{P_2^{L-}}^{P_2^L} V_2(P_1^{H+}, t)dt \\
\geq \; & \varepsilon\Delta_2^H V_2(P_1^{L-}, P_2^L) - \delta\Delta_2^L V_2(P_1^{H+}, P_2^L) \geq \left[\varepsilon\Delta_2^H - \delta\Delta_2^L\right] V_2(P_1^{H-}, P_2^{H-}) \geq 0 \qquad (16)
\end{aligned}$$

where the first inequality is strict when $\Delta_2^H, \Delta_2^L > 0$.

Now use (15) and (16) to sign the change in the sender's expected payoff from any mean improving inframodular swap:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L-}, P_2^{H+}) - V(P_1^L, P_2^H)\right] + \delta\left[V(P_1^{H+}, P_2^{L-}) - V(P_1^H, P_2^L)\right] \\
= \; & \varepsilon\left[V(P_1^{L-}, P_2^H) - V(P_1^L, P_2^H)\right] + \delta\left[V(P_1^{H+}, P_2^L) - V(P_1^H, P_2^L)\right] \\
& + \varepsilon\left[V(P_1^{L-}, P_2^{H+}) - V(P_1^{L-}, P_2^H)\right] + \delta\left[V(P_1^{H+}, P_2^{L-}) - V(P_1^{H+}, P_2^L)\right] \geq 0
\end{aligned}$$

where the inequality is strict when either strict when $\Delta_1^H, \Delta_1^L > 0$ or $\Delta_2^H, \Delta_2^L > 0$.

Step 2: Positive Expansions. Using in order: V bi-convex (with $P_1^H > P_1^L$), then V SPM (with $P_2^H > P_2^L$), and finally $\delta\Delta_1^H \geq \varepsilon\Delta_1^L$ (with V non-decreasing) we discover:

$$\begin{aligned}
& \varepsilon\left[V(P_1^{L-}, P_2^L) - V(P_1^L, P_2^L)\right] + \delta\left[V(P_1^{H+}, P_2^H) - V(P_1^H, P_2^H)\right] \\
= \; & \delta \int_{P_1^H}^{P_1^{H+}} V_1(s, P_2^H)ds - \varepsilon \int_{P_1^{L-}}^{P_1^L} V_1(s, P_2^L)ds \\
\geq \; & \delta\Delta_1^H V_1(P_1^H, P_2^H) - \varepsilon\Delta_1^L V_1(P_1^H, P_2^L) \geq \left[\delta\Delta_1^H - \varepsilon\Delta_1^L\right] V_1(P_1^H, P_2^L) \geq 0 \qquad (17)
\end{aligned}$$

where the first inequality is strict when $\Delta_1^H, \Delta_1^L > 0$.

Similarly, using in order: V bi-convex (with $P_2^H > P_2^L$), then V SPM (with $P_1^{H+} >$

P_1^{L-}), and finally $\delta\Delta_2^H \geq \varepsilon\Delta_2^L$ (with V non-decreasing) we discover:

$$\begin{aligned} & \varepsilon\left[V(P_1^{L-},P_2^{L-})-V(P_1^{L-},P_2^{L})\right]+\delta\left[V(P_1^{H+},P_2^{H+})-V(P_1^{H+},P_2^{H})\right] \\ = & \ \delta\int_{P_2^H}^{P_2^{H+}} V_2(P_1^{H+},t)dt-\varepsilon\int_{P_2^{L-}}^{P_2^{L}} V_2(P_1^{L-},t)dt \\ \geq & \ \delta\Delta_2^H V_2(P_1^{H+},P_2^H)-\varepsilon\Delta_2^L V_2(P_1^{L-},P_2^H) \geq \left[\delta\Delta_2^H-\varepsilon\Delta_2^L\right]V_2(P_1^{L-},P_2^H) \geq 0 \end{aligned} \quad (18)$$

where the first inequality is strict when $\Delta_2^H, \Delta_2^L > 0$.

Now use (17) and (18) to sign the change in the sender's expected payoff from any mean improving inframodular swap:

$$\begin{aligned} & \varepsilon\left[V(P_1^{L-},P_2^{L-})-V(P_1^{L},P_2^{L})\right]+\delta\left[V(P_1^{H+},P_2^{H+})-V(P_1^{H},P_2^{H})\right] \\ = & \ \varepsilon\left[V(P_1^{L-},P_2^{L})-V(P_1^{L},P_2^{L})\right]+\delta\left[V(P_1^{H+},P_2^{H})-V(P_1^{H},P_2^{H})\right] \\ & +\varepsilon\left[V(P_1^{L-},P_2^{L-})-V(P_1^{L-},P_2^{L})\right]+\delta\left[V(P_1^{H+},P_2^{H+})-V(P_1^{H+},P_2^{H})\right] \geq 0 \end{aligned}$$

where the inequality is strict when either strict when $\Delta_1^H, \Delta_1^L > 0$ or $\Delta_2^H, \Delta_2^L > 0$. □

6.3 Strong Polarization with Sufficient Disagreement

We say that signal pair j, j' is a *polarizing pair* if $(P_1^j - P_1^{j'})(P_2^j - P_2^{j'}) \leq 0$. A message service is *strongly polarizing* if it induces at least two distinct posterior pairs $P^j \neq P^{j'}$ and every pair of signals is a polarizing pair.[9] Equivalently, the receiver's are in perfect ordinal disagreement about how to interpret the messages. That is, the signal that makes receiver 1 most confident that the state is θ_1 makes receiver 2 least confident that the state is θ_1, and the signal that makes receiver 1 second most confident the state is θ_1 makes receiver 2 second most confident the state is θ_1, etc.

Notice that V bi-concave and submodular, implies that $V(P,P)$ is concave in P, precisely the assumption needed to induce the sender to conceal all information in the standard common prior model. But in our model with sufficient ex ante disagreement about the payoff irrelevant state, these assumptions yield a very different conclusion:

Proposition 3. *Assume private communication. If V is bi-concave and SBM (at least one strict), then the optimal message service must be strongly polarizing.*

Proof: Toward a contradiction, assume an optimal message service induces a pair of signals j, j' with $(P_1^j - P_1^{j'})(P_2^j - P_2^{j'}) > 0$ WLOG with $P_1^j < P_1^{j'}$ and $P_2^j < P_2^{j'}$.

[9] Requiring two distinct posteriors rules out perfectly concealing message services, which are certainly not polarizing, but would otherwise trivially satisfy the definition.

Step 1: Payoff Improving Contractions for $\Pi_1^{j'}/\Pi_1^j \geq \Pi_2^{j'}/\Pi_2^j$. Fix the probabilities Π_i^j and $\Pi_i^{j'}$ and posterior beliefs P_2^j and $P_2^{j'}$, while increasing P_1^j by $\Delta_1^j > 0$ and decreasing $P_1^{j'}$ by $\Delta_1^{j'} > 0$ without changing receiver 1's expected belief, i.e.:

$$\Pi_1^j \Delta_1^j - \Pi_1^{j'} \Delta_1^{j'} = 0 \tag{19}$$

By construction, this change has no impact on the sender's expectation of receiver 2's posterior. Using (19) and the assumed inequality $\Pi_1^{j'}/\Pi_1^j \geq \Pi_2^{j'}/\Pi_2^j$, we can sign the change in the senders expectation of receiver 1's posterior $\Delta E_s[P_1]$.

$$\begin{aligned}
(19) \quad &\Rightarrow \quad \Delta_1^j - \frac{\Pi_1^{j'}}{\Pi_1^j}\Delta_1^{j'} = 0 \quad \Rightarrow \quad \Delta_1^j - \frac{\Pi_2^{j'}}{\Pi_2^j}\Delta_1^{j'} > 0 \\
&\Rightarrow \quad \Delta E_s[P_1] = (1-\alpha)\left(\Pi_2^j \Delta_1^j - \Pi_2^{j'} \Delta_1^{j'}\right) > 0
\end{aligned}$$

Altogether, we have constructed a mean improving positive contraction, which must increase the sender's payoff by Lemma 3. Further, this change is feasible since it respects the Bayesian constraint (7) for each receiver.

Step 2: Payoff Improving Contractions for $\Pi_1^{j'}/\Pi_1^j < \Pi_2^{j'}/\Pi_2^j$. Fix the probabilities Π_i^j and $\Pi_i^{j'}$ and posterior beliefs P_1^j and $P_1^{j'}$, while increasing P_2^j by $\Delta_2^j > 0$ and decreasing $P_2^{j'}$ by $\Delta_2^{j'} > 0$ without changing receiver 1's expected belief, i.e.:

$$\Pi_2^j \Delta_2^j \Pi_2^{j'} \Delta_2^{j'} = 0 \tag{20}$$

By construction, this change has no impact on the sender's expectation of receiver 1's posterior. Using (20) and the assumed inequality $\Pi_1^{j'}/\Pi_1^j < \Pi_2^{j'}/\Pi_2^j$, we can sign the change in the senders expectation of receiver 2's posterior $\Delta E_s[P_2]$.

$$\begin{aligned}
(20) \quad &\Rightarrow \quad \Delta_2^j - \frac{\Pi_2^{j'}}{\Pi_2^j}\Delta_2^{j'} = 0 \quad \Rightarrow \quad +\Delta_2^j - \frac{\Pi_1^{j'}}{\Pi_1^j}\Delta_2^{j'} \geq 0 \\
&\Rightarrow \quad \Delta E_s[P_2] = \alpha\left(\Pi_1^j \Delta_2^j - \Pi_1^{j'} \Delta_2^{j'}\right) \geq 0
\end{aligned}$$

Altogether, we have constructed a negative contraction, which must increase the sender's payoff by Lemma 3. Further, this change is feasible since it respects the Bayesian constraint (7) for each receiver.

For the more general case of $q_1 < q_2$, we can still construct payoff improving contractions as in each step. However, exactly one of the two steps will violate the consistency constraint (8). As a result, this case requires additional restrictions on V.

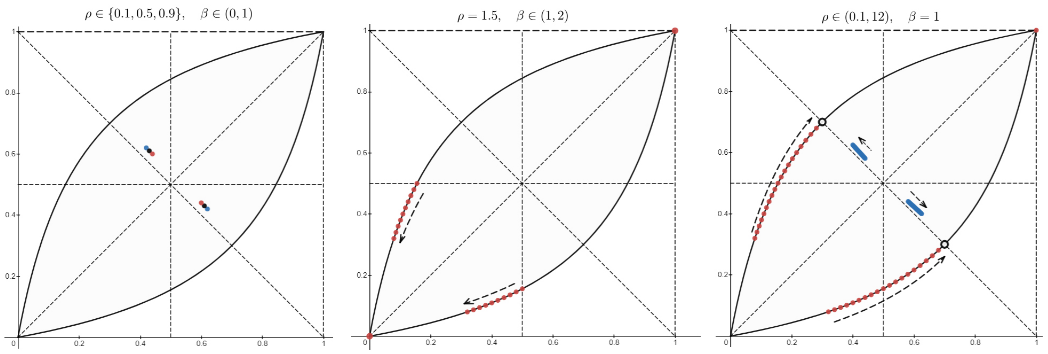

Figure 4: **Changes in Posteriors for CES Values** (21)**.** Left: When $(\beta, \rho) \in (0,1)$, the solution is strongly polarizing, independent of β, and polarization increases as ρ rises from 0.25 (red) to 0.5 (black) to 0.75 (blue). Middle: fixing $\rho = 1.5$, the optimal posterior distribution is supported on $((x,y),(y,x),(1,1))$. When $\beta < \beta^* \approx 1.4$, (x,y) falls in β, and full revelation ($x = y = 0$) is optimal for all $\beta > \beta^*$. Right: fixing $\beta = 1$, the solution is strongly polarizing for $\rho < 1$ and polarization rises in ρ (blue). When $\rho \in [1,6]$ the support is $((x,y),(y,x),(1,1))$ (red), and, as ρ rises, the pair $((x,y),(y,x))$ converges along the lens to the intersection of the lens and $P_2 = 1 - P_1$, which remains the solution for all $\rho > 6$.

7 Comparative Statics in Examples

In this section we consider the comparative statics of polarization. Throughout, we use the following cardinal measure of polarization for any distribution over posterior beliefs (P^j, Π^j) :

$$\Phi \equiv \sum_j \left(\alpha \Pi_1^j + (1-\alpha)\Pi_2^j\right) |P_1^j - P_2^j|$$

That is, the expected distance between the receiver's posterior beliefs from the sender's perspective. Results for alternative natural measures of polarization are qualitatively similar (e.g., covariance between receiver posteriors).

7.1 Changes in the Sender's Value Function.

Consider the following sender value function:

$$V(P_1, P_2) = (P_1^\rho + P_2^\rho)^{\frac{\beta}{\rho}} \tag{21}$$

This value is bi-concave when $\beta, \rho \leq 1$ and bi-convex when $\beta, \rho \geq 1$. It is strictly SPM when $\beta > \rho$ and strictly SBM when $\beta < \rho$. For now we fix beliefs $q_1 = 0.3, q_s = 0.5, q_2 = 0.7$, and $p = 0.5$, and see how the solution changes as we vary β and ρ.

Figure 4 illustrates how the support of the posterior belief distribution (P_1^j, P_2^j) varies in β and ρ. In all graphs, the black lens shaped curve demarcates the region of possible posteriors (i.e. pairs of posteriors (P_1, P_2) that can be generated by some message service given the receivers prior beliefs [10]). When V is bi-concave, i.e., $(\beta, \rho) \in (0,1)^2$, strongly polarizing, binary messages are optimal, and polarization increases in ρ. When $\rho > 1$, there is a cutoff β^*, such that full revelation is optimal for all $\beta > \beta^*$. And when $\beta < \beta^*$, the support of the posterior belief distribution involves three distinct pairs $((x,y),(y,x),(1,1))$. That is, optimal message services involve three signals: one reveals that the state is θ_1 and the other two signals form a polarizing pair. In the knife-edged case of $\beta = 1$, strongly polarizing binary signals are optimal for $\rho < 1$, trinary signals are optimal for $\rho \in [1,6)$. As ρ rises from 1 to 6, the revealing signal becomes less likely and the distance between the polarized outcomes increases; and thus, polarization rises.[11] For $\rho \geq 6$, a strongly polarizing binary message service is again optimal. Note that in all cases, polarization Φ rises in ρ (the elasticity of substitution between receiver posterior beliefs).

7.2 Changes in Receiver Beliefs.

Now we fix the sender's value function and vary receiver beliefs, relaxing the consistency constraint (8). Specifically, we fix $p = 1/2, q_s = 1/2$ and $q = q_2 = 1 - q_1$ and vary q on $[0.54, 0.99]$. Let V be the bi-concave and SBM function given by (21) with $\beta = 1/2$ and $\rho = 3/4$. In this case, optimal message services are binary and strongly polarizing. As the ex ante disagreement between the receivers increases, polarization rises, as does $\mathrm{E}_s[P_1, P_2]$. In the limit, the support of the belief distribution converges to $((1,1/2),(1/2,1))$ (Figure 5, left), illustrating Proposition 3.

Toward an intuition for this case, recall that we fixed the common prior in the payoff relevant state to $p = 1/2$. In order to achieve the posterior beliefs pictured in the left hand graph of Figure 5, the sender uses the following binary message service:

$$\begin{array}{ccc} & \pi^1 : & \\ & \theta_0 & \theta_1 \\ \omega_0 & 0 & \varepsilon \\ \omega_1 & 1 & 1-\varepsilon \end{array} \qquad\qquad \begin{array}{ccc} & \pi^2 : & \\ & \theta_0 & \theta_1 \\ \omega_0 & 1 & 1-\varepsilon \\ \omega_1 & 0 & \varepsilon \end{array} \tag{22}$$

When ex ante disagreement on state ω is very high and ε is very low, each of these signals is almost perfectly revealing for one of the receivers, while simultaneously providing almost no information to the other receiver. As $q_1 \to 0$ and $q_2 \to 1$, the sender chooses $\varepsilon \to 0$ and the posterior belief distribution converges to $(P^1, P^2) = ((1,1/2),(1/2,1))$

[10] See the proof in Appendix C

[11] For fixed $\rho \in (1,6)$, the optimal distribution places less weight on $(1,1)$ as β rises, and for β sufficiently large the optimal message service is binary and strongly polarizing (case not pictured).

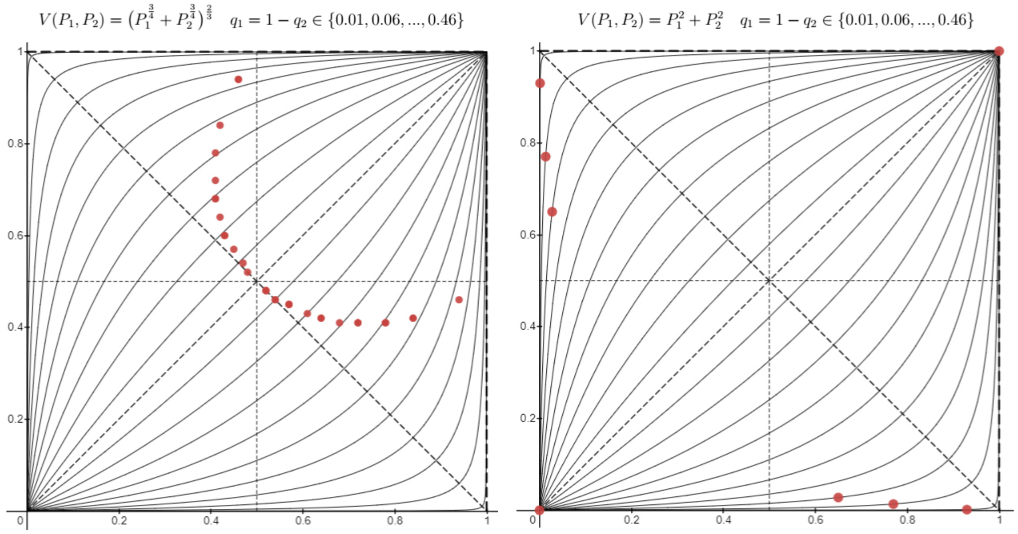

Figure 5: **Changes in Posteriors in Ex Ante Disagreement.** Left: When the CES value function (21) is bi-concave and SBM the optimal message service is binary and strongly polarizing. As we increase the receivers ex ante disagreement on the payoff irrelevant state, polarization rises. In the private communication limit, the support converges to $((1, 1/2), (1/2, 1))$. On the Right, the sender's value is the bi-convex and modular function $V(P_1, P_2) = P_1^2 + P_2^2$. Perfect revelation is optimal when receiver disagreement is not too high, but past a threshold level of disagreement, posterior beliefs have trinary support $((x, y), (y, x), (1, 1))$ with $x \to 1$ and $y \to 0$ in the private communication limit.

with chances $(\Pi^1, \Pi^2) = ((0, 1), (1, 0))$.

For a bi-convex value function case, let $V(P_1, P_2) = P_1^2 + P_2^2$. Since V is convex, the sender prefers spread posteriors, all else equal. She could induce polarization, thereby increasing $\mathrm{E}_s[(P_1, P_2)]$, but this would necessitate a lower spread in beliefs in order to comply with the consistency constraint (8). Thus, there is a tradeoff between higher mean posterior beliefs and lower spread. When q_1 and q_2 are close together, achieving higher mean belief entails sacrificing too much spread, and the sender chooses a perfectly revealing message service, as in the standard common prior case. But when q_1 and q_2 are sufficiently far apart, the sender chooses a partially revealing message service; namely, a distribution with trinary support $((x, y), (y, x), (1, 1))$. As $q_2 = 1 - q_1 \to 1$, $x \to 1$ and $y \to 0$, and the distribution converges to the private communication limit in Proposition 2.

8 Conclusion

We have extended the canonical Bayesian persuasion game in order to explore incentives and abilities of the sender to polarize her audience. In our model, it is no longer true that sender conceals information when her payoff is concave in beliefs and fully reveals the state when her payoff is convex in beliefs. When the sender's payoff function is differentiable and increasing, even slight disagreement on the payoff irrelevant dimension is enough to guarantee that sender will use an informative signal. As a consequence, the sender strictly benefits from increased disagreement on the payoff irrelevant prior for such payoff functions.

We characterized solutions for extreme disagreement when the sender's payoff is biconvex in the receivers' posteriors beliefs and also when the payoff is biconcave and submodular. In the biconvex case, two out of three signals induce extreme polarization. In the biconcave and submodular case, the two receivers perfectly disagree about the ordinal interpretation of signals.

We illustrated via numerical examples that extreme ex ante disagreement is not necessary for polarization. These simulation results also support our conjecture that when the sender's prior belief in payoff irrelevant state is close enough to $(q_1 + q_2)/2$, the sender induces strong polarization given any biconcave and submodular payoff function.

In addition, we are currently working on monotone comparative statics. Namely, polarization orders on posterior distributions and orders on the set of payoff functions, such that a monotone change in the payoff function induces a monotone change in posterior polarization.

References

Alesina, A. F., A. Miano, and S. Stantcheva (2020): "The Polarization of Reality," Working Paper 26675, National Bureau of Economic Research.

Andreoni, J., and T. Mylovanov (2012): "Diverging Opinions," *American Economic Journal: Microeconomics*, 4(1), 209–32.

Aumann, R. J. (1976): "Agreeing to Disagree," *Ann. Statist.*, 4(6), 1236–1239.

——— (1995): *Repeated Games with Incomplete Information*, vol. 1 of *MIT Press Books*. The MIT Press.

Baliga, S., E. Hanany, and P. Klibanoff (2013): "Polarization and Ambiguity," *American Economic Review*, 103(7), 3071–83.

Benoit, J. P., and J. Dubra (2019): "Apparent Bias: What Does Attitude Polarization Show?," *International Economic Review*, 60, 1675 –1703.

Crawford, V., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–51.

Fryer Jr, R., P. Harms, and M. Jackson (2019): "Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization," *Journal of the European Economic Association*, 17(5), 1470–1501.

Jeong, D. (2019): "Using cheap talk to polarize or unify a group of decision makers," *Journal of Economic Theory*, 180(C), 50–80.

Jern, A., K. Chang, and C. Kemp (2014): "Belief polarization is not always irrational.," *Psychological review*, 121 2, 206–24.

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

Loh, I., and G. Phelan (2019): "Dimensionality and Disagreement: Asymptotic Belief Divergence in Response to Common Information," *International Economic Review*, 60(4), 1861–1876.

Lord, C. G., and L. Ross (1979): "Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence," *Journal of Personality and Social Psychology*, pp. 2098–2109.

Morris, S. (1994): "Trade with Heterogeneous Prior Beliefs and Asymmetric Information," *Econometrica*, 62(6), 1327–1347.

——— (1995): "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 11(2), 227–253.

Muller, A., and M. Scarsini (2012): "Fear of Loss, Inframodularity, and Transfers," *Journal of Economic Theory*, 147, 1490–1500.

Pkhakadze, N. (2021): "Polarizing Cheap Talk," Discussion paper, Georgetown University.

Rabin, M., and J. L. Schrag (1999): "First Impressions Matter: A Model of Confirmatory Bias*," *The Quarterly Journal of Economics*, 114(1), 37–82.

Ricardo, A., and O. Camara (2016): "Bayesian Persuasion with Heterogeneous Priors," *Journal of Economic Theory*, 165, 672–706.

Van den Steen, E. (2010a): "Culture Clash: The Costs and Benefits of Homogeneity," *Management Science*, 56(10), 1718–1738.

——— (2010b): "Disagreement and the Allocation of Control," *Journal of Law, Economics, and Organization*, 26(2), 385–426.

A A Second Reformulation

The reformulation in Lemma 1 is useful for understanding the impact of receiver disagreement on the sender's optimal *payoff*, but requires $q_1 \neq q_2$. In this section we derive a second reformulation that can be applied directly to the case $q_1 = q_2$. Toward this reformulation, let $s = ((1-p)(1-q_s), p(1-q_s), (1-p)q, pq)$ denote the sender's prior belief over the four states $((\theta_0, \omega_0), (\theta_1, \omega_0), (\theta_0, \omega_1), (\theta_1, \omega_1))$, and let $S^j = (S_1^j, S_2^j, S_3^j, S_4^j)$ be the sender's posterior belief vector over these four states given some signal j. Let $\Delta(S)$ be the set of distributions on the 3-simplex: $\{(S_1+S_2+S_3) \geq 0 | S_2+S_3+S_4 \leq 1\}$ with positive probability on four realizations.

Define receiver i's *relative prior bias* $\varrho_i \equiv \frac{(1-q_i)q_s}{(1-q_s)q_i} - 1$. Thus, ϱ_i is positive when $q_s > q_i$ and negative when $q_s < q_i$. We can then express receiver posterior beliefs on the payoff relevant state θ as a function of the sender's posterior beliefs on the two dimensional state, and reformulate the sender's maximization as a choice over her beliefs, subject to the usual martingale restriction.

Lemma 5. *If a signal generates sender posterior beliefs S then receiver beliefs are:*

$$\mathcal{P}_i(S) \equiv Pr(\theta = \theta_1 | S) = \frac{1 - S_1 - S_3 + S_2 \varrho_i}{1 + \varrho_i(S_1 + S_2)}$$

And sender maximization (3) *is equivalent to:*

$$V^* = \max_{\delta \in \Delta(s)} E_\delta[V(\mathcal{P}_1(S), \mathcal{P}_2(S))] \quad s.t. \; E_\delta[S] = s$$

When $\varrho_i = 0$, we have $\mathcal{P}_i(S) = 1 - S_1 - S_2$, that is, the receiver and sender share the same beliefs on the payoff relevant state. Thus, if $\varrho_1 = \varrho_2 = 0$, Lemma 5 reduces to the standard formulation in Kamenica and Gentzkow (2011).

Proof: The proof relies on the theorem from Ricardo and Camara (2016). Assume Ω is finite set of states and $a(w), b(w)$ are priors of two Bayesian agents where $a, b \in int(\Delta(\Omega))$. Assume that both agents observe the same signal realization.
If posterior of agent with prior $b(\omega)$ is $B(\omega)$ then the posterior of agent with prior $a(\omega)$ is given by:

$$A(\omega) = B(\omega) \cdot \frac{a(\omega)}{b(\omega)} \cdot \frac{1}{\Sigma_{\alpha \in \Omega} B(\alpha) \frac{a(\alpha)}{b(\alpha)}} = \frac{B(\omega) \cdot \frac{a(\omega)}{b(\omega)}}{E_B(\frac{a}{b})} = \frac{B(\omega) \cdot \ell(\omega)}{E_B(\ell)}$$

Where by $\ell(\omega)$ is denoted $\frac{a(\omega)}{b(\omega)}$ and E_B is expectation w.r.t distribution B.
$\ell(\omega)$ is how much more likely agent with prior a thinks that the state is ω in comparison to agent with prior b.

Let $\Omega = \{\omega_1, ..., \omega_n\}$ then a, b can be represented as vectors in $(0,1)^n$. $a = (a_1, ..., a_n)$ and $b = (b_1, ..., b_n)$ s.t. $a_i = a(\omega_i)$ and $b_i = b(\omega_i)$

Similarly denote by $B = (B_1, ...B_n)$ and $A = (A_1, ...A_n)$ posteriors $B(\omega)$ and $A(\omega)$.

Assume agents observe some arbitrary signal realization with likelihood $x = (x_1, ..., x_n) \in [0, 1]^n$ meaning that the probability of this signal realization in state ω_i is x_i. Then $\forall i$

$$B_i = \frac{b_i x_i}{\Sigma_j b_j x_j}$$

$$A_i = \frac{a_i x_i}{\Sigma_j a_j x_j}$$

Since $a_i \neq 0$ and $b_i \neq 0$ we can divide the equations on each other whenever $x_i \neq 0$. If $x_i = 0$ then both $A_i = 0$ and $B_i = 0$ and the statement of the lemma holds trivially. So assume $x_i \neq 0$ and divide the second equation on the first one:

$$\frac{A_i}{B_i} = \frac{a_i}{b_i} \cdot \frac{\Sigma_j b_j x_j}{\Sigma_k a_k x_k} = \ell_i \cdot \frac{1}{\frac{\Sigma_k a_k x_k \cdot \frac{b_k}{b_k}}{\Sigma_j b_j x_j}} = \ell_i \cdot \frac{1}{\Sigma_k \ell_k \cdot \frac{x_k b_k}{\Sigma_j b_j x_j}} = \ell_i \cdot \frac{1}{\Sigma_k \ell_k B_k} \Rightarrow$$

$$\Rightarrow P_i = \frac{B_i \ell_i}{\Sigma_k \ell_k B_k} \Rightarrow A(\omega) = \frac{\ell(\omega) B(\omega)}{\Sigma_{\alpha \in \Omega} B(\alpha) \ell(\alpha)} = \frac{B(\omega) \ell(\omega)}{E_B(\ell)}$$

Given this representation we can express the posteriors of two receivers in terms of sender's posteriors and the transform sender's objective function to a function of her own posterior as it is given the lemma 5. This finishes the proof.

Corollary 2. *If V is an upper semicontinuous and V^* is achieved by some message service, then it could be also achieved by a message service with at most 4 signals. In other words, sender's expected payoff space is spanned by message services of at most 4 signals.*

Proof: Corollary follows from Lemma 5 and a general result which states that Caratheodory number of the hypograph of a an upper semicontinuous function of k variable is at most k. To see this, assume that $(p^*, d^*) \in (0, 1)^n \times (0, 1)^{n(k-1)}$ is the solution of the following maximization problem:

$$\begin{array}{ll} \max\limits_{p,d} & \Sigma p_i V(d_i) \\ \text{subject to} & \Sigma p_i d_i = d_0 \end{array}$$

Where $p \in \cup_j (0, 1)^j$, $d_i \in (0, 1)^{k-1}$ for $i \in \{0, 1, ...\}$ and $V : [0, 1]^{k-1} \mapsto R$ is upper semicontinuous. Let $\Sigma_{i=1}^n p_i^* v(d_i^*) = v^*$, then:

$$\exists p \in [0, 1]^k \textit{ and } \exists d \in [0, 1]^{k(k-1)} \textit{ s.t.}$$

$$\Sigma_{i=1}^k p_i v(d_i) = v^*$$

If V is such that $V(d_0) = V^*$ then $p = (1, 0, ...0)$ and $d = (d_0, e, ...e)$ or if $n < k$

$p = (p_1^*, ...p_n^*, 0, ..., 0)$ and $d = (d_1^*, ...d_n^*, e, ..., e)$ where $e \in [0,1]^{k-1}$ is arbitrary.
More interesting is the case when $n > k$. Let $Co(V)$ be the concavification of V, then from Aumann (1995), or from Kamenica and Gentzkow (2011), we know that (d_0, V^*) should be located on the graph of V.
(d_0, v^*) is in the convex hull of hypograph of $Co(V)$, which is the subset of R^k. By Caratheodory's theorem (d_0, V^*) lies in k-simplex with vertices in the hypograph of V. There can be many such k-simplexes, but all of them should be the subsets of convex hull of hypograph of V i.e. subset of hypograph of $Co(V)$.
The proof is based on the following claim: *there should at least on k-simplex for which (d_0, V^*) isn't an interior point.* Otherwise it would be an interior point of hypograph of $Co(V)$ and thus cannot be located on the graph of $Co(V)$. From this follows that (d_0, V^*) is not an interior point for at least one of the k-simplexes meaning that (d_0, V^*) is located on one out of the k "borders" of k-simplex, which are $(k-1)$-simplexes. Consequently (d_0, V^*) is convex combination of k points from hypograph of V.
The fact that vertices of $(k-1)$-simplex cannot be from interior of hypograph of V is trivially contradicting that V^* is the maximum, so vertices of $(k-1)$-simplex must be located on the graph of V, meaning that (d_0, V^*) can be attained by convex combination of k points of type $(d_i, V(d_i))$. This ends our proof. Q.E.D.

B Omitted Proofs

B.1 Proof of Lemma 1

Assume an optimal message service for the maximization (3) has n realizations, $\{m_1, ..., m_n\}$; and thus, can be represented as n probability matrices.

$$\begin{array}{ccc} & \theta_0 & \theta_1 \\ \omega_0 & b_j & a_j \\ \omega_1 & B_j & A_j \end{array}$$

where
$$\sum_{i=1}^{n} A_j = \sum_{i=1}^{n} B_j = \sum_{i=1}^{n} a_j = \sum_{i=1}^{n} b_j = 1 \tag{23}$$

We need to show that for any distribution over posterior beliefs that satisfies constraints (7) and (8), we can find nonnegative A_j, B_j, a_j, b_j satisfying (23) and vice versa, i.e., for every nonnegative A_j, B_j, a_j, b_j, satisfying (23) we can find P_1, P_2, Π_1, Π_2 satisfying (7) and (8). We will show the first direction, the second direction follows easily from reversing each step.

Let us denote by P_i^j posterior of player i after observing m_j and by Π_i^j the probability of realization m_j for player i. And let A_j, B_j, a_j, b_j satisfy

$$P_i^j = \frac{pq_iA_j + p(1-q_i)a_j}{pq_iA_j + p(1-q_i)a_j + (1-p)q_iB_j + (1-p)(1-q_i)b_j}$$

$$\Pi_i^j = pq_iA_j + p(1-q_i)a_j + (1-p)q_iB_j + (1-p)(1-q_i)b_j$$

Now let us trace back A_j, B_j, a_j, b_j from Π_i^j and P_i^j.

$$P_i^j = \frac{pq_iA_j + p(1-q_i)a_j}{pq_iA_j + p(1-q_i)a_j + (1-p)q_iB_j + (1-p)(1-q_i)b_j} = \frac{pq_iA_j + p(1-q_i)a_j}{\Pi_i^j} \Leftrightarrow$$

$$\Leftrightarrow q_iA_j + (1-q_i)a_j = \frac{P_i^j\Pi_i^j}{p} \; for \; i = 1,2 \; and \; j = 1,...,n$$

$$\Updownarrow$$

$$A_j = \frac{P_2^j\Pi_2^j(1-q_1) - P_1^j\Pi_1^j(1-q_2)}{p(q_2-q_1)}$$

$$a_j = \frac{P_1^j\Pi_1^jq_2 - P_2^j\Pi_2^jq_1}{p(q_2-q_1)}$$

Similarly for B_j and b_j we get:

$$P_i^j = 1 - \frac{(1-p)q_iB_j + (1-p)(1-q_i)b_j}{pq_iA_j + p(1-q_i)a_j + (1-p)q_iB_j + (1-p)(1-q_i)b_j} \Leftrightarrow$$

$$\Leftrightarrow q_iB_j + (1-q_i)b_j = \frac{(1-P_i^j)\Pi_i^j}{1-p} \; for \; i = 1,2 \; and \; j = 1,...,n$$

$$\Updownarrow$$

$$B_j = \frac{(1-P_2^j)\Pi_2^j(1-q_1) - (1-P_1^j)\Pi_1^j(1-q_2)}{(1-p)(q_2-q_1)}$$

$$b_j = \frac{(1-P_1^j)\Pi_1^jq_2 - (1-P_2^j)\Pi_2^jq_1}{(1-p)(q_2-q_1)}$$

Let us check that (23) holds:

$$\sum_{i=1}^n A_j = \sum_{i=1}^n \frac{P_2^j\Pi_2^j(1-q_1) - P_1^j\Pi_1^j(1-q_2)}{p(q_2-q_1)} = \frac{\sum_{i=1}^n P_2^j\Pi_2^j(1-q_1) - \sum_{i=1}^n P_1^j\Pi_1^j(1-q_2)}{p(q_2-q_1)} =$$

$$= \frac{p(1-q_1) - p(1-q_2)}{p(q_2-q_1)} = 1$$

In a similar fashion we get that $\sum_{i=1}^n B_j = \sum_{i=1}^n a_j = \sum_{i=1}^n b_j = 1$. Finally, we show that if (8) holds, then A_j, a_j, B_j, b_j are nonnegative.

Given P_i^j, Π_i^j we found A_j, a_j, B_j, b_j which sum up to 1. In order this A_j, a_j, B_j, b_j to be legitimate likelihoods we need to have $A_j, a_j, B_j, b_j \in [0,1]$ and since they sum up to 1, each of them is nonnegative will be enough. Since $1 \geq q_2 > q_1 \geq 0$

$$A_j \geq 0 \Leftrightarrow P_2^j\Pi_2^j(1-q_1) - P_1^j\Pi_1^j(1-q_2) \geq 0 \Leftrightarrow \frac{P_1^j\Pi_1^j}{P_2^j\Pi_2^j} \leq \frac{1-q_1}{1-q_2}$$

$$a_j \geq 0 \Leftrightarrow P_1^j\Pi_1^j q_2 - P_2^j\Pi_2^j q_1 \geq 0 \Leftrightarrow \frac{P_1^j\Pi_1^j}{P_2^j\Pi_2^j} \geq \frac{q_1}{q_2}$$

Combining the last two we get:

$$\Big(A_j \geq 0 \wedge a_j \geq 0\Big) \Leftrightarrow \frac{q_1}{q_2} \leq \frac{P_1^j\Pi_1^j}{P_2^j\Pi_2^j} \leq \frac{1-q_1}{1-q_2}$$

As for B_j and b_j, note that

$$B_j \geq 0 \Leftrightarrow (1-P_2^j)\Pi_2^j(1-q_1) - (1-P_1^j)\Pi_1^j(1-q_2) \geq 0 \Leftrightarrow \frac{(1-P_1^j)\Pi_1^j}{(1-P_2^j)\Pi_2^j} \leq \frac{1-q_1}{1-q_2}$$

$$b_j \geq 0 \Leftrightarrow (1-P_1^j)\Pi_1^j q_2 - (1-P_2^j)\Pi_2^j q_1 \geq 0 \Leftrightarrow \frac{(1-P_1^j)\Pi_1^j}{(1-P_2^j)\Pi_2^j} \geq \frac{q_1}{q_2}$$

Combining the last two we get:

$$\Big(B_j \geq 0 \wedge b_j \geq 0\Big) \Leftrightarrow \frac{q_1}{q_2} \leq \frac{(1-P_1^j)\Pi_1^j}{(1-P_2^j)\Pi_2^j} \leq \frac{1-q_1}{1-q_2}$$

So we get that if Π_i^j and P_i^j are s.t. (7) and (8) hold, we can construct likelihood matrixes which generate these Π_i^j and P_i^j and this construction is only possible when (7) and (8) hold. As for equality of objective functions note that:

For the sender probability of m_j outcome is:

$$\Pi_s^j = pq_sA_j + p(1-q_s)a_j + (1-p)qB_j + (1-p)(1-q_s)b_j =$$

$$= p(\alpha q_1 + (1-\alpha)q_2)A_j + p(1-(\alpha q_1 + (1-\alpha)q_2))a_j +$$

$$+(1-p)(\alpha q_1 + (1-\alpha)q_2)B_j + (1-p)(1-(\alpha q_1 + (1-\alpha)q_2))b_j =$$

$$= \alpha\Pi_1^j + (1-\alpha)\Pi_2^j$$

So objective function for both maximization problems are the same. This finishes the proof.

C The Set of Possible Posteriors

In this section we'll derive the condition which must a pair of posterior beliefs to satisfy in order it to be attainable when prior beliefs are given by $p, q_1, \text{ and } q_2$.
The resulted set is a lens shaped subset of $[0,1]^2$, which is an area between two hyperbolic functions and could be written in the following way:

$$\Big\{(P_1, P_2)\Big| \frac{P_1}{\frac{(1-q_1)q_2}{(1-q_2)q_1}(1-P_1)+P_1} \leq P_2 \leq \frac{P_1}{\frac{(1-q_2)q_1}{(1-q_1)q_2}(1-P_1)+P_1}\Big\} \bigcap \Big[0,1\Big]^2$$

We'll say that posterior $(P_1, P_2) \in [0,1]^2$ is attainable for prior believes $p, q_1, \text{ and } q_2$, if there exist an signal which induces posterior belief of receiver i to be equal to P_i. Let's denote the set of possible posteriors by $B(p, q_1, q_2)$.

claim 1 $\forall p, p' \neq 0$, $B(p, q_1, q_2) = B(p', q_1, q_2)$
Proof: Take arbitrary $(P_1, P_2) \in B(p, q_1, q_2)$, by definition $\exists x, y, z, t \geq 0$ s.t. after observing event (signal realization) with likelihood

$$\begin{array}{ccc} & \theta_0 & \theta_1 \\ \omega_0 & t & z \\ \omega_1 & y & x \end{array}$$

Receiver i's belief becomes P_i, i.e.:

$$\frac{pq_1x + p(1-q_1)z}{pq_1x + p(1-q_1)z + (1-p)q_1y + (1-p)(1-q_1)t} = P_1$$

$$\frac{pq_2x + p(1-q_2)z}{pq_2x + p(1-q_2)z + (1-p)q_2y + (1-p)(1-q_2)t} = P_2$$

Now consider receivers with priors: $p', q_1, \text{ and } q_2$ and event with likelihood:

$$\begin{array}{ccc} & \theta_0 & \theta_1 \\ \omega_0 & t' & z' \\ \omega_1 & y' & x' \end{array}$$

where $x' = x \cdot \frac{p}{p'}$, $y' = y \cdot \frac{1-p}{1-p'}$, $z' = z \cdot \frac{p}{p'}$ and $t' = t \cdot \frac{1-p}{1-p'}$. Posterior of receiver i after observing this event will be:

$$\frac{p'q_1x' + p'(1-q_i)z'}{p'q_ix' + p(1-q_i)z' + (1-p)q_iy' + (1-p)(1-q_i)t} =$$

$$\frac{p'q_1x \cdot \frac{p}{p'} + p'(1-q_i)z \cdot \frac{p}{p'}}{p'q_ix \cdot \frac{p}{p'} + p(1-q_i)z \cdot \frac{p}{p'} + (1-p)q_iy \cdot \frac{1-p}{1-p'} + (1-p)(1-q_i)t \cdot \frac{1-p}{1-p'}} =$$

$$= \frac{pq_ix + p(1-q_i)z}{pq_ix + p(1-q_i)z + (1-p)q_iy + (1-p)(1-q_i)t} = P_i \Rightarrow$$

$$\Rightarrow (P_1, P_2) \in B(p', q_1, q_2) \Rightarrow B(p, q_1, q_2) \subseteq B(p', q_1, q_2)$$

$B(p, q_1, q_2) \supseteq B(p', q_1, q_2)$ is analogous. Q.E.D.

Claim 1 gives us opportunity to simplify notation and the further analyzes, from now instead of $B(p, q_1, q_2)$, we'll use $B(q_1, q_2)$ and assume that $p = 0.5$.

Note that after observing events E_1 and E_2 with the following likelihoods:

E_1	θ_0	θ_1
ω_0	t	z
ω_1	y	x

E_2	θ_0	θ_1
ω_0	αt	αz
ω_1	αy	αx

posteriors are the same $\forall \alpha > 0$. So we can assume that $t = 1$ ($\alpha = \frac{1}{t}$, or any of x, y, z can be 1). So

$$(P_1, P_2) \in B(q_1, q_2) \Leftrightarrow (P_1, P_2) \in B(0.5, q_1, q_2) \Leftrightarrow \exists (x, y, z, t) > (0,0,0,0) \ s.t.$$

$$\frac{\frac{1}{2}q_1x + \frac{1}{2}(1-q_1)z}{\frac{1}{2}q_1x + \frac{1}{2}(1-q_1)z + (1-\frac{1}{2})q_1y + (1-\frac{1}{2})(1-q_1)t} = P_1$$

$$\frac{\frac{1}{2}q_2x + \frac{1}{2}(1-q_2)z}{\frac{1}{2}q_2x + \frac{1}{2}(1-q_2)z + (1-\frac{1}{2})q_2y + (1-\frac{1}{2})(1-q_2)t} = P_2$$

$$\Updownarrow$$

$$\exists (x, y, z) \geq (0,0,0) \ s.t.$$

$$\frac{q_1x + (1-q_1)z}{q_1x + (1-q_1)z + q_1y + (1-q_1)} = P_1 \bigwedge \frac{q_2x + (1-q_2)z}{q_2x + (1-q_2)z + q_2y + (1-q_2)} = P_2$$

So we are looking for P_1 and P_2 s.t. this system of equations with 2 equations and 3 unknowns has non-negative solution. Let us solve for x and z.

$$\frac{q_1x + (1-q_1)z}{q_1x + (1-q_1)z + q_1y + (1-q_1)} = P_1 \bigwedge \frac{q_2x + (1-q_2)z}{q_2x + (1-q_2)z + q_2y + (1-q_2)} = P_2$$

$$\Updownarrow$$

$$\begin{cases} (1-P_1)(q_1x + (1-q_1)z) = P_1(q_1y + (1-q_1)) \\ (1-P_2)(q_2x + (1-q_2)z) = P_2(q_2y + (1-q_2)) \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} q_1x + (1-q_1)z = \frac{P_1}{1-P_1}(q_1y + 1 - q_1) \\ q_2x + (1-q_2)z = \frac{P_2}{1-P_2}(q_2y + 1 - q_2) \end{cases}$$

$$\Updownarrow$$

Introduce notations $\frac{P_i}{1-P_i} \equiv A_i$ and $\frac{1-q_i}{q_i} \equiv Q_i$

$$\begin{cases} x + Q_1 z = A_1(y + Q_1) \\ x + Q_2 z = A_2(y + Q_2) \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} \frac{1}{Q_1}x + z = A_1(\frac{1}{Q_1}y + 1) \\ \frac{1}{Q_2}x + z = A_2(\frac{1}{Q_2}y + 1) \end{cases}$$

$$\Updownarrow$$

$$z = \frac{A_1 - A_2}{Q_1 - Q_2}y + \frac{Q_1A_1 - Q_2A_2}{Q_1 - Q_2} \bigwedge x = \frac{A_1Q_2 - A_2Q_1}{Q_2 - Q_1}y + \frac{Q_1Q_2(A_1 - A_2)}{Q_2 - Q_1}$$

WLOG assume that $q_2 > q_1$ (If $q_1 = q_2$ then $B(q_1, q_2) = \{(P_1, P_2)|P_1 = P_2\} \cap [0, 1]^2$) and then $Q_1 > Q_2$. Also since $P_i > 0$, $A_1 > A_2 \Leftrightarrow P_1 > P_2$.

We are looking for (P_1, P_2) s.t. $\exists y \geq 0$ s.t. $x \geq 0$ and $z \geq 0$

$$z = \frac{A_1(y + Q_1) - A_2(y + Q_2)}{Q_1 - Q_2} \geq 0 \Leftrightarrow A_1(y + Q_1) - A_2(y + Q_2) \geq 0$$

$$x = \frac{A_2Q_1(y + Q_2) - A_1Q_2(y + Q_1)}{Q_1 - Q_2} \geq 0 \Leftrightarrow A_2Q_1(y + Q_2) - A_1Q_2(y + Q_1) \geq 0$$

$$\Updownarrow$$

$$\begin{cases} A_1(y + Q_1) \geq A_2(y + Q_2) \\ A_2Q_1(y + Q_2) \geq A_1Q_2(y + Q_1) \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} y(A_1 - A_2) \geq A_2Q_2 - A_1Q_1 \\ y(A_2Q_1 - A_1Q_2) \geq Q_1Q_2(A_1 - A_2) \end{cases}$$

- Case 1, $A_1 = A_2$: In this case it's easy to check that both inequalities above hold for $\forall y \geq 0$. So all $A_1 = A_2$ work for us.
  $A_1 = A_2 \Leftrightarrow P_1 = P_2 \Rightarrow \{(P_1, P_2)|P_1 = P_2\} \cap [0, 1]^2 \subseteq B(q_1, q_2)$
- Case 2, $A_1 > A_2$: Since $Q_1 > Q_2$, $A_2Q_2 - A_1Q_1 < 0$ and $y(A_1 - A_2) \geq 0$, $\forall y \geq 0$, So the first inequality holds for all non-negative y-s.
  Right hand side of the second inequality is positive, so if want that inequality holds for some non-negative y it's necessary and sufficient that: $A_2Q_1 - A_1Q_2 > 0$.
  Equivalently
  $$\frac{Q_1}{Q_2} > \frac{A_1}{A_2} > 1$$

- Case 3 $A_1 < A_2$: In this case the second inequality holds for all non-negative y-s and in order the first inequality to have non-negative solution it's necessary and sufficient that: $A_2Q_2 - A_1Q_1 < 0$. Equivalently:

$$\frac{Q_1}{Q_2} > \frac{A_2}{A_1} > 1$$

If we combine this three cases and substitute back q_i-s and P_i-s we get the following condition.

$\forall p > 0$ and $\forall q_1 \leq q_2 \in (0, 1)$, if prior beliefs are given by p, q_1, q_2, then the set of possible posteriors is given by:

$$B(q_1, q_2) = \Big\{(P_1, P_2)\Big| \frac{P_1}{\frac{(1-q_1)q_2}{(1-q_2)q_1}(1-P_1)+P_1} \leq P_2 \leq \frac{P_1}{\frac{(1-q_2)q_1}{(1-q_1)q_2}(1-P_1)+P_1}\Big\} \bigcap \Big[0, 1\Big]^2$$