Bayesian Persuasion with Private Information*

Andrew Kosenko†

February 5, 2018

Abstract

We study a model of communication and Bayesian persuasion between a sender who is privately informed and has state independent preferences, and a receiver who has preferences that depend on the unknown state. In a model with two states of the world, over the interesting range of parameters, the equilibria can be pooling or separating, but a particular novel refinement forces the pooling to be on the most informative information structure in all but one case. We also study two extensions - a model with more information structures as well as a model where the state of the world is non-dichotomous, and show that analogous results emerge.

JEL Classification: D82, D83, C72.

Keywords: persuasion, communication, information provision, information design.

*I would like to thank Jose Scheinkman and Joseph Stiglitz for useful discussions, Yeon-Koo Che and the participants in the economic theory colloquium at Columbia for discussion and comments, and most of all, I am very grateful to my advisor, Navin Kartik, for his continued support, help and encouragement.

†Columbia University, Department of Economics; ak2912@columbia.edu

1 Introduction

When can one interested party persuade another interested party of something? This question is of major economic interest, since persuasion, broadly construed, is crucial to many economic activities. As pointed out by Taneva (2016), there are basically two ways of persuading any decision maker to take an action - one is by providing the appropriate incentives (this, of course, is the subject of mechanism design), and the other by providing appropriately designed information. Indeed, design of informational environments as well as their effect on strategic interaction has been the subject of much study for at least fifty years in economics and is continuing to yield new results. In the present work we focus on a more specific question - namely when the party that is doing the persuading is inherently interested in a specific outcome, and in addition, has some private information about the problem. In a setting of mutual uncertainty about the true state of the world, the problem information design with private information on one side has a number of interesting features not to mention the myriad possible applications. In this work we model this situation, explore the equilibria and their properties (welfare and comparative statics), and show that a particular equilibrium refinement nearly always selects the equilibria with the most information revelation (in a sense to be made precise below).

This particular setup is motivated by two important leading examples - the trial process where a prosecuting attorney[1] is trying to persuade a jury and a judge of the guilt of a defendant, and the setting of drug approval where a pharmaceutical company is trying to persuade the Federal Drug Administration of the value of a new drug. In both settings the party that is trying to convince the other party of something may (and in fact, typically, does) have private information about the true state of the world. In the case of the prosecution attorney, this may be something that the defendant had privately indicated to the counsel, and in the case of the pharmaceutical company this may be some internal data or the views of scientists employed by the company. But in both cases the persuading party has to conduct a publicly visible experiment (a public court trial or a drug clinical trial, exhibiting the testing protocol in advance) that may reveal something hitherto unknown to either party. A key assumption that we make is this: the evidence, whether it is favorable (in an appropriate sense) to the prosecutor or drug company, or not, from such an experiment cannot be concealed; if that were possible the setup would be related to the literature on verifiable disclosure ("hard information") initiated by Milgrom (1981) and Grossman (1981).

The setting is one of a communication game with elements of persuasive signaling. There is a single sender and a single receiver. There is an unknown state of the world (going along with one of the analogies from above, we may describe the state space as $\Omega = \{\text{Innocent}, \text{Guilty}\}$). Neither the sender nor the receiver know the true state, and the have a commonly known prior belief about the true state. To justify this assumption we appeal to the fact that in the two main applications described it is, indeed, satisfied[2]. The sender obtains a private, imperfectly informative

[1] One could just as well think of the case of a defense attorney - they key elements of the environment will be preserved.

[2] In fact, in the drug approval example nobody at all knows the true state, and in the court example only the defendant knows the true state - but she is not able to signal it credibly.

signal about the state of the world, and armed with that knowledge[3] has to choose an information structure that will generate a signal that is again imperfectly informative of the state. The receiver then has to take an action, based on the prior belief, the choice of information structure as well as the realization of the signal, that will affect the payoffs of both parties. This kind of a situation is ubiquitous in real life, and certainly deserves much attention.

The game has elements of several modeling devices; first of all there's the signaling element - different types of sender have different types corresponding to their privately known subjective beliefs. However, these types do not enter into either party's preferences - that's the cheap talk (Crawford and Sobel (1982)) element. Finally there is the element of Bayesian persuasion (see Kamenica and Gentzkow (2011)) since all types of sender can choose all possible information structures (in other words, the set of available information structures does not depend on the sender's type), but cannot fully control the signal that will be realized according to that information structure.

The main difference of this model is that the heterogeneity of the sender is not about who she is (such as, for example, in basic signaling[4] and screening models) or what she does (such as in models involving moral hazard), but purely in what she *knows*. The preferences of the different types of sender are identical (so that, in particular, there is no single-crossing or analogous assumption on the preferences). Their type doesn't enter their payoff function; in fact, not even their action enters their payoff directly - it does so only through the effect it has on the action of the receiver. This assumption is at odds with much of the literature on the economics of information; it is intended to capture the intuition that there is nothing intrinsically different in the different types of senders and to isolate the effect of private information on outcomes.

Although this setting is certainly rather permissive, we do not consider a number of important issues. In particular, there is no "competition in persuasion" here - there are no informational contests between the prosecution side and the defense side or competing drug firms designing trials about each other's candidate drugs (although this is an interesting possibility that is explored in Gentzkow and Kamenica (2017a) an Gentzkow and Kamenica (2017b)). In similar settings (but without private information) it has been shown in previous work (Gentzkow and Kamenica (2017a)) that competition typically, though not always, improves overall welfare and generates "more" information. Furthermore, in the present setting, the "persuader" is providing information about the relevant state of the world; another interesting possibility is signaling about one's private information. For example, the prosecuting attorney could provide verifiable evidence not of the form "the investigation revealed certain facts", but rather, verifiable evidence of the form "I think the defendant is guilty because of the following:...". We also assume that the receiver does not have commitment power; namely he cannot commit to doing something (say, taking an action that is very bad for the sender) unless he observes the choice of a very informative experiment; do-

[3]Note that at that point, the beliefs of the sender and receiver about the state of the world will no longer agree in general, so that one may think of this situation as analogous to starting with heterogeneous priors; see Alonso and Camara (2016c).

[4]With the exception of cheap talk models, which do have this feature.

ing so would not be subgame-perfect on the part of the receiver. Finally, we assume that choosing different information structures has the same cost which we set to zero.

In the present paper we also make an additional assumption that signals that reveal the state fully are either unavailable, or prohibitively costly. In any realistic setting this is true. We will show that this assumption, along with others, is important in the kinds of equilibria that can arise; notably, this assumption will reverse some of the previous results about coexistence of different equilibria and their welfare properties. This is among the primary contributions of this work.

The rest of the paper is organized as follows. In the next section, we discuss the literature and place the present model in context. Section 3 describes in detail the setting, the basic model and derives the main results; we fully characterize the equilibria of the model and show the ways in which the outcomes are different from existing work. Section 4 extends the model. Section 5 concludes.

2 Relationship to Existing Literature

This work is in the spirit of the celebrated approach of Kamenica and Gentzkow (2011) ("KG" from here onward) on so-called "Bayesian persuasion". Among the key methodological contributions of that work is the fact that they show that the payoff of the sender can be written as a function of the posterior of the receiver; they also identify conditions under which the sender "benefits from persuasion", utilizing a "concavification" technique introduced in Aumann and Maschler (1995).

Hedlund (2016) is the most closely related work in this area; he works with a very similar model but he assumes that the sender has a very rich set of experiments available; in particular, an experiment that fully reveals the payoff-relevant state is available. He also places a number of other assumptions, such as continuity, compactness and strict monotonicity on relevant elements of the model. We present an independently conceived and developed model but acknowledge having benefitted from seeing his approach. This work provides context to his results in the sense that we consider a simpler model where we can explore the role of particular assumptions and show the importance of these features for equilibrium welfare. In particular, we consider experiments where a fully revealing signal is not available; this assumption seems more realistic in applications and creates an additional level of difficulty in analysis that is not present in Hedlund (2016). In addition, we show that dropping any of the assumptions in that work produces a model the equilibria of which closely resemble the equilibria we find in the present work.

Perez-Richet (2014) considers a related model where the type of the sender is identified with the state of the world; there the sender is, in general, not restricted in the choice of information structures. He characterizes equilibria (of which there are many) and applies several refinements to show that in general, predictive power of equilibria is weak, but refinements lead to the selection of the high-type optimal outcome. His model is a very special case of the model presented here.

Degan and Li (2015) study the interplay between the prior belief of a receiver and the precision of (costly) communication by the sender; they show that all plausible equilibria must involve pooling. In addition, they compare results under two different strategic environments - one where the sender can commit to a policy before learning any private information, and one without such commitment, and again derive welfare properties that are dependent on the prior belief. Akin to Perez-Richet (2014), they identify the type of sender with the state of the world.

Alonso and Camara (2016a) show that in general, the sender can not benefit from becoming an expert (i.e. from learning some private information about the state). This result also hinges on the existence of a fully revealing experiment, an assumption that we do not make in this work; in our setting the sender may or may not benefit from persuasion.

Other related work includes Rayo and Segal (2010), who show that a sender typically benefits from partial information disclosure. Gill and Sgroi (2012) study an interesting and related model in which a sender can commit to a public test about her type. Alonso and Camara (2016c) present a similar models where the sender and receiver have different, but commonly known priors about the state of the world. The model in this paper can be seen as a case of a model where the sender and receiver also have different priors, but the receiver does not know the prior of the sender. In addition, Alonso and Camara (2016c) endow their senders with state-dependent utility functions. In related work, there are also many current projects extending this sort of informative persuasion to models of voting (Arieli and Babichenko (2016), Alonso and Camara (2016b)).

3 Model

3.1 Basic setup (2 states, 2 types of sender, 2 experiments, 2 signals, 2 actions for receiver)

To fix ideas and generate intuition we first study a simplified model, and then extend the results. Let us consider a strategic communication game between a sender (she) and receiver (he), where the sender (S) has private information. In contrast with Perez-Richet (2014), the private information of the sender is not about who she is (her type), but about what she knows about the state of the world. In Perez-Richet (2014)'s work the sender is perfectly informed about her type (which is also the state of the world). In this setup this is not true. The sender is imperfectly informed about the state of the world. Consequently, the receiver (R) will have beliefs about both the type of the sender and the state of the world.

There is a single state of the world, $\omega \in \Omega = \{\omega_H, \omega_L\}$, unknown to both parties with a commonly known prior probability of $\omega = \omega_H$ equal to $\pi \in (0,1)$. The sender can can be one of two types: $\theta \in \Theta = \{\theta_H, \theta_L\}$. The sender's type is private information to her. The type structure is generated as follows:

$$\mathbb{P}(\theta = \theta_H | \omega = \omega_H) = \mathbb{P}(\theta = \theta_L | \omega = \omega_L) = \xi \tag{1}$$

and

$$\mathbb{P}(\theta = \theta_H | \omega = \omega_L) = \mathbb{P}(\theta = \theta_L | \omega = \omega_H) = 1 - \xi \tag{2}$$

for $\xi \geq \frac{1}{2}$

This is the key feature distinguishing this model from others - the private information of the sender is not about her preferences (as in Perez-Richet (2014), and more generally, in mechanism design by an informed principal), but about the state of nature. In this sense the sender is more informed than the receiver. The sender chooses an *experiment* - a complete conditional distribution of *signals* given states[5]; all experiments have the same cost, which we set to zero[6]. The choice of the experiment and the realization of the signal are observed by both the sender and the receiver. For now the sender is constrained to choose among two experiments; the available experiments are:

$$\Pi_H = \begin{array}{c c} & \begin{array}{cc} \omega_H & \omega_L \end{array} \\ \begin{array}{c} \sigma_H \\ \sigma_L \end{array} & \begin{pmatrix} \rho_H & 1-\rho_H \\ 1-\rho_H & \rho_H \end{pmatrix} \end{array}$$

and

$$\Pi_L = \begin{array}{c c} & \begin{array}{cc} \omega_H & \omega_L \end{array} \\ \begin{array}{c} \sigma_H \\ \sigma_L \end{array} & \begin{pmatrix} \rho_L & 1-\rho_L \\ 1-\rho_L & \rho_L \end{pmatrix} \end{array}$$

The entries in the matrices represent the probabilities of observing a signal (only two are available: σ_H and σ_L) conditional on the state. We also assume that $\rho_H > \rho_L$, and say that Π_H is *more informative* than Π_L[7]. The available actions for the receiver are $a \in \{a_H, a_L\}$.

3.2 Preferences

The sender has state-independent preferences, always preferring action a_H. The receiver, on the other hand, prefers to take the high action in the high state and the low action in the low state. To fix ideas, suppose that $u^S(a_H) = 1$, $u^S(a_L) = 0$, and the receiver has preferences given by $u^R(a, \omega)$. We will state some basic results without specifying and explicit functional form, and then make more assumptions to derive meaningful results. Importantly, there is no single-crossing assumption on the primitives in this model. Rather, a similar kind of feature is derived endogenously.

One can also consider $a \in A$ with A a compact subset of $\mathbb{R}$, and preferences of the form (for the sender) $u^S(\omega, a) = \tilde{u}^S(a)$ with $\tilde{u}^S$ a strictly increasing function, and (for the receiver) $u^R(\omega, a) = \tilde{u}^R(\omega, a)$ with $\tilde{u}^R$ having increasing differences in the two arguments, as does Hedlund (2016) in his work. It turns out that this specification has substantially different implications for equilibria and equilibrium selection. In addition, in applications (and certainly in the motivating

[5]The are many terms for what we are calling an "experiment" in the literature; in particular, "information structure" and "signal".

[6]As opposed to Degan and Li (2015) who posit costly signals.

[7]It so happens that all experiments in this section are also ranked by Blackwell's criterion but we do not use this fact.

examples discussed above) it seems more natural to work with a discrete action space.

3.3 Timing

The timing of the game is as follows:

1. Nature chooses the state, ω.
2. Given the choice of the state, Nature generates a type for the sender according to the distribution above.
3. The sender privately observes the type and chooses an experiment.
4. The choice of the experiment is publicly observed. The receiver forms interim beliefs about the state.
5. The signal realization from the experiment is publicly observed. The receiver forms posterior beliefs about the state.
6. The receiver takes an action and payoffs are realized.

3.4 Analysis

It will be convenient to let $p(\theta) = \mathbb{P}(\Pi = \Pi_H|\theta)$ be the (possibly mixed) strategy of the sender and $q(\Pi, \sigma) = \mathbb{P}(a = a_H|\Pi, \sigma)$ that of the receiver. Denoting by "hats" the observed realizations of random variables and action choices, let $\mu(\hat{\omega}|\hat{\Pi}) = \mathbb{P}\left(\omega = \hat{\omega}|\Pi = \hat{\Pi}\right)$ be the *interim* (i.e. before observing the realization of the signal from the experiment) belief of the receiver about the state of the world, given the observed experiment., and write $\mu(\hat{\Pi}) = \mathbb{P}(\omega = \omega_H|\Pi = \hat{\Pi})$. Let $\beta(\omega_H|\Pi, \sigma)$ be the *posterior* belief of the receiver that the state is high conditional on observing Π and σ, given interim beliefs μ. Thus, $\beta(\hat{\Pi}, \hat{\sigma}) = \mathbb{P}\left(\omega = \omega_H|\Pi = \hat{\Pi}, \sigma = \hat{\sigma}, \mu\right)$. It is notable that here what matters are the beliefs of the receiver about the payoff-relevant random variable (the state of the world), as opposed to beliefs about the type of the sender, as in the vast majority of the literature. However, one does need to have beliefs about the type of the sender to be able to compute overall beliefs in a reasonable way; to that end let $\nu(\theta|\Pi) = \mathbb{P}(\theta|\Pi)$ be the beliefs of the receiver about the type of the sender, conditional on observing an experiment Π. These beliefs are an equilibrium object, and necessary to compute the interim beliefs μ; we will however, suppress the dependence of μ on ν to economize on notation in hopes that the exposition will be clear enough.

Let $v(\Pi, \theta, q) \triangleq \mathbb{E}\left(u^S(a)|\Pi, \theta, q\right)$ be the expected value of announcing experiment Π for a sender of type θ. For example,

$$v(\Pi_H, \theta_H, q) = \rho_H \mathbb{P}(\omega_H|\theta_H) q(\Pi_H, \sigma_H) + (1 - \rho_H)\mathbb{P}(\omega_H|\theta_H) q(\Pi_H, \sigma_L) + \tag{3}$$

$$+ (1 - \rho_H)\mathbb{P}(\omega_L|\theta_H) q(\Pi_H, \sigma_H) + \rho_H \mathbb{P}(\omega_L|\theta_H) q(\Pi_H, \sigma_L) \tag{4}$$

One can compute $v(\Pi_H, \theta_L, q), v(\Pi_L, \theta_H, q)$ and $v(\Pi_L, \theta_L, q)$ in a similar fashion. Also let

$$v(p(\theta), \theta, q) \triangleq p(\theta)v(\Pi_H, \theta, q) + (1 - p(\theta))v(\Pi_L, \theta, q) \tag{5}$$

In any equilibrium[8], the receiver must be best-responding given his beliefs, or :

$$a^*(\Pi, \sigma) \in \arg\max_{\Delta\{a_H, a_L\}} u^R(a, \omega_H)\beta(\Pi, \sigma) + u^R(a, \omega_L)(1 - \beta(\Pi, \sigma)) \tag{6}$$

and $q^*(\Pi, \sigma) = \mathbb{P}(a^* = a_H | \Pi, \sigma)$.

Following the notation in the literature, let $\hat{v}(\Pi_i, \mu, \theta_j) \triangleq \mathbb{E}_{\sigma,a}(u^S(a)|\Pi_i, \mu)$ denote the expected value of choosing an experiment Π_i for type θ_j when the receiver's interim beliefs are exactly μ. Thus,

$$\begin{aligned}\hat{v}(\Pi_i, \mu, \theta_j) \triangleq \rho_i \left[\mathbb{P}(\omega_H|\theta_j)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_j)\mathbb{1}_{|\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_i)\left[\mathbb{P}(\omega_H|\theta_j)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_j)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}}\right]\end{aligned} \tag{7}$$

The function $\hat{v}$ is piecewise linear in μ and continuous in the choice of the experiment (equivalently, in ρ_i).

3.5 Perfect Bayesian equilibria

For concreteness, and to allow explicit calculation of equilibria, for the rest of this section we will focus on a particular form for the preferences of the receiver; namely, suppose that $u^R(\omega_H, a_H) = 1, u^R(\omega_H, a_L) = -1, u^R(\omega_L, a_L) = 1, u^R(\omega_L, a_H) = -1$. The symmetry in the payoffs is special, but doesn't affect the qualitative properties of equilibria.

As a first step we can see what happens in the absence of asymmetric information - that is, when both the sender and the receiver can observe the type of the sender. In that case the interim belief of the receiver is based on the observed type of the sender (instead of the observed choice of experiment): $\mu(\theta) = \mathbb{P}(\omega = \omega_H|\theta)$ and the strategy of receiver is modified accordingly to $q(\theta, \sigma) = \mathbb{P}(a = a_H|\theta, \sigma)$. The decision of the sender is then reduced to choosing the experiment that yields the higher expected utility. In other words,

$$\forall\theta, p(\theta) = 1 \iff v(\Pi_H, \theta, q) > v(\Pi_L, \theta, q) \tag{8}$$

and $p(\theta) = 0$ otherwise (ties are impossible given the different parameters and the specification of the sender's utility). Observe that this situation is identical to to the model described in KG (and all the insights therein apply), except that the sender is constrained to choose among only two experiments.

From now assume that the type of sender is privately known only to the sender. As a first observation one can note that in any equilibrium we must have $p(\theta_H) \geq p(\theta_L)$; otherwise one

[8] We discuss existence below.

would get an immediate contradiction.

Definition 1. *A weak perfect Bayesian equilibrium with tie-breaking (or "equilibrium", for brevity) is a four-tuple $(p(\theta), a^*(\Pi, \sigma), \mu, \beta)$ that satisfy the following conditions:*

1. *Sequential Rationality:*

$$\forall \theta, p(\theta) \in \textit{arg max } v(\Pi, \theta, q) \textit{ and } a^*(\Pi, \sigma) \in \arg\max \sum_{\omega} u(a, \omega) \beta(\omega|\Pi, \sigma) \tag{9}$$

2. *Consistency: μ and β are computed using Bayes rule whenever possible, taking into account the strategy of the sender as well as equilibrium interim beliefs about the type of sender.*

3. *Tie-breaking: whenever $\beta(\Pi, \sigma) = \frac{1}{2}, a^*(\Pi, \sigma) = a_H$.*

The moniker "weak" in this definition is meant to draw attention to the fact that off the equilibrium path beliefs of the sender are unrestricted, a fact that will come in useful in supporting some equilibria. The first two parts of the definition are standard. We augment the definition with a tie-breaking rule (the third requirement) to facilitate and simplify the exposition. The rule requires that whenever the receiver is indifferent between two actions, he always chooses the one preferred by the sender[9]. A more substantive reason to focus on this particular tie-breaking rule is that this makes the value function of the sender upper-semicontinuous, and so by an extended version of the Weierstrass theorem, there will exist an experiment maximizing it. This will be crucial when we consider more inclusive sets of experiments.

For the question of existence[10] of equilibria one can appeal to the fact that this is a finite extensive game, and as such, has a trembling-hand perfect equilibrium (Selten (1975) and Osborne and Rubinstein (1994), their Corollary 253.2), and therefore, has a sequential equilibrium (Kreps and Wilson (1982), and therefore has a wPBE, since these equilibrium concepts are nested.

As usual, in evaluating the observed signal the receiver uses a conjecture of the sender's strategy, correct in equilibrium. Note once again that in contrast to Hedlund (2016), in the present model there is no experiment that fully discloses the state of the world. If it was available, and the sender were to choose it, then the sender's payoffs would be independent of the receiver's interim belief (rendering the entire "persuasion" point moot); such an experiment would also provide uniform type-specific lower bounds on payoffs for the sender, since that would be a deviation that would always be available. The fact that this is not available makes the analysis more difficult, but also more interesting. The preference specification in the present model allows us to get around the difficulty and derive analogous results without relying on the existence of a perfectly revealing experiment.

[9]It is common in the literature to focus on "sender-preferred" equilibria; we do not make the same assumption, but "bias" out equilibria in the same direction

[10]Even though we explicitly construct an equilibrium, and hence they certainly exist, it is useful to have a result for more general settings.

In what follows we will focus on the interesting range of parameters $\{\pi, \xi, \rho_H, \rho_L\} \in (0,1) \times \left(\frac{1}{2}, 1\right)^3$, where the receiver takes different actions after different signals[11]. To that end, let

Definition 2 (Nontrivial equilibria)**.** *An equilibrium is said to be fully nontrivial (or just nontrivial) in pure strategies if $a^*(\Pi_i, \sigma_H) = a_H$, $a^*(\Pi_i, \sigma_L) = a_L$, for both $\Pi_i \in \{\Pi_H, \Pi_L\}$; that is, the receiver follows the signal in these equilibria.*

Definition 3 (P-nontrivial equilibria)**.** *An equilibrium is said to be partially nontrivial (or p-nontrivial) in pure strategies if $a^*(\Pi_i, \sigma_H) = a_H$ and $a^*(\Pi_i, \sigma_L) = a_L$, for one $\Pi_i \in \{\Pi_H, \Pi_L\}$, but not both. That is, the receiver follows the signal realization after observing one but not the other experiment.*

Other possibilities may arise: one can define nontrivial and p-nontrivial equilibria mixed strategies analogously. However, either kind of non-trivial equilibria in mixed strategies are ruled out by the tie-breaking assumption made earlier; as a consequence we do not consider such equilibria. It is immediate that if an equilibrium is nontrivial, it is also p-nontrivial, but not vice versa. From now on we will focus only on (p-)nontrivial equilibria; this amounts to placing restrictions on the four parameters that we will be explicit about when convenient. This clearly doesn't cover all possible equilibria for all possible parameters, but it does focus on the "interesting" equilibria. The following straightforward propositions serve to narrow down the set of possible equilibria.

Proposition 1. *Suppose that an equilibrium is p-nontrivial. Then in such an equilibrium both types of sender use the same pure strategy.*

Proof. The fact that both types of sender must use a pure strategy follows from the fact that in any p-nontrivial equilibrium choosing one experiment strictly dominates choosing another, regardless of the beliefs of the sender or the interim beliefs of the receiver receiver. The fact that that pure strategy must be the same for both types also follows from the same observation. □

Proposition 2. *Suppose that an equilibrium is fully nontrivial. In such an equilibrium it must be the case that each type chooses the experiment that maximizes the probability of generating a "high" signal, without regard to the effect of the choice of experiment on in the interim belief. Moreover, each type of sender uses a pure strategy.*

Proof. Take a fully nontrivial equilibrium. In any such equilibrium the receiver follows the observed signal with probability one, for any experiment. Therefore it must be the case that each type of sender is best-responding by simply evaluating the expected probability of the "high" signal (noting that the utility of a low action, which would result from a low signal, is zero, and thus the probability of the low signal can be ignored), and is choosing whichever experiment delivers the higher probability, ignoring the problem of signaling one's type by choice of experiment, since

[11] There always exist parameters (and payoffs) such that regardless of the choice of experiment and signal realization, the receiver always takes the same action, or ignores the signal and takes an action based purely on the chosen experiment. We do not focus on these equilibria. Also note that the issue of nontrivial equilibria does not arise in a model with a compact action space.

for any such choice, the interim belief would still result in a fully nontrivial equilibrium, by assumption. Ties are impossible due to the different precision of experiments and different sender beliefs, hence the focus on pure strategies. □

The above two propositions taken together eliminate the possibility of mixing for the sender. The following propositions state all possible equilibria; they are supported, as is standard, by beliefs that assign probability one to off-path deviations coming from the low type of sender. Incentive compatibility can be proven by directly computing utilities on and off the equilibrium path, and verifying best responses, using Bayes rule whenever possible. We omit the tedious but straightforward computations. For convenience, for any variable $x \in (0,1)$ denote by $\tilde{x}$ the ratio $\frac{x}{1-x}$.

Proposition 3. *There is a unique separating equilibrium where $p(\theta_H) = 1, p(\theta_L) = 0$. This equilibrium exists as long as $\{\pi, \xi, \rho_H, \rho_L\}$ satisfy equations the following restrictions: $\pi \leq \xi, \pi + \xi > 1, \tilde{\pi}\tilde{\rho_H}\tilde{\xi} > 1, \tilde{\rho_H} > \tilde{\pi}\tilde{\xi}, \tilde{\pi}\tilde{\rho_L} > \tilde{\xi}, \tilde{\rho_L}\tilde{\xi} > \tilde{\pi}$. Denote this equilibrium by "SEP".*

Intuitively, in this equilibrium the low type of sender prefers to "confuse" the receiver by sending a sufficiently uninformative signal. We now turn to classifying pooling equilibria.

Proposition 4. *There is a continuum of fully nontrivial pooling equilibria where $p(\theta_H) = p(\theta_L) = 1$. These equilibria exist as long as $\pi + \xi \geq 1, \pi \geq \xi, \tilde{\pi}\tilde{\rho_H} \geq 1, \rho_H > \pi, \tilde{\pi}\tilde{\rho_L} \geq \tilde{\xi}, \tilde{\rho_L}\tilde{\xi} > \tilde{\pi}$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_L) \in [\mathbb{P}(\omega_H|\theta_L), \rho_L)$. Denote this kind of equilibria by "FNT-H".*

Proposition 5. *There is a continuum of fully nontrivial pooling equilibria where $p(\theta_H) = p(\theta_L) = 0$. These equilibria exist as long as $\pi + \xi \leq 1, \pi \leq \xi, \tilde{\pi}\tilde{\rho_H}\tilde{\xi} \geq 1, \rho_L > \pi, \tilde{\rho_L} > \tilde{\xi}\tilde{\pi}, \tilde{\rho_L}\tilde{\pi} \geq 1$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_H) \in [\mathbb{P}(\omega_H|\theta_L), \rho_H)$. Denote this kind of equilibria by "FNT-L".*

Proposition 6. *There is a continuum of p-nontrivial pooling equilibria where $p(\theta_H) = p(\theta_L) = 1, a^*(\Pi_L, \sigma) = a_L$, for $\sigma = \sigma_H, \sigma_L$, and $a^*(\Pi_H, \sigma_H) = a_H, a^*(\Pi_H, \sigma_L) = a_L$.. These equilibria exist as long as $\tilde{\xi} > \tilde{\rho_L}\tilde{\pi}, \rho_H > \pi$, and $\pi + \rho_H \geq 1$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_L) \in [\mathbb{P}(\omega_H|\theta_L), 1-\rho_L)$. Denote this kind of equilibria by "PNT-HL(a_L)"*[12].

Proposition 7. *There is a continuum of p-nontrivial pooling equilibria where $p(\theta_H) = p(\theta_L) = 1, a^*(\Pi_H, \sigma) = a_H$, for $\sigma = \sigma_H, \sigma_L$ and $a^*(\Pi_L, \sigma_H) = a_H, a^*(\Pi_L, \sigma_L) = a_L$. These equilibria exist as long as $\tilde{\rho_L}\tilde{\pi} \geq \tilde{\xi}, \rho_H \geq \pi, \tilde{\pi} < \tilde{\xi}\tilde{\rho_L}$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_L) \in [\mathbb{P}(\omega_H|\theta_L), \rho_L)$. Denote this kind of equilibria by "PNT-HH(a_H)".*

Proposition 8. *There is a continuum of p-nontrivial pooling equilibria where $p(\theta_H) = p(\theta_L) = 0, a^*(\Pi_L, \sigma_H) = a_H, a^*(\Pi_L, \sigma_L) = a_L$ and $a^*(\Pi_H, \sigma) = a_L$, for $\sigma = \sigma_H, \sigma_L$. These equilibria exist as long as $\rho_L > \pi$, $\rho_L + \pi \geq 1$ and $\tilde{\rho_H}\tilde{\pi} < \tilde{\xi}$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_H) \in [\mathbb{P}(\omega_H|\theta_L), 1-\rho_H)$. Denote this kind of equilibria by "PNT-LH(a_L)".*

[12] For any PNT equilibrium, the notation "PNT-XY(a_i)" equilibrium denotes the fact that the senders pool on experiment X, and the receiver takes the same action after observing experiment Y, for $X, Y = H, L, a_i \in \{a_H, a_L\}$.

Proposition 9. *There is a continuum of p-nontrivial pooling equilibria where $p(\theta_H) = p(\theta_L) = 0, a^*(\Pi_L, \sigma) = a_H$, for $\sigma = \sigma_H, \sigma_L$ and $a^*(\Pi_H, \sigma_H) = a_H, a^*(\Pi_H, \sigma_L) = a_L$. These equilibria exist as long as $\tilde{\rho_H}\tilde{\pi} \geq \tilde{\xi}, \rho_L \leq \pi, \tilde{\pi} \leq \tilde{\xi}\tilde{\rho_H}$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_H) \in [\mathbb{P}(\omega_H|\theta_L), 1 - \rho_L)$. Denote this kind of equilibria by "PNT-LL(a_H)".*

These are all the equilibria of this game[13]. The following proposition, which can be verified by direct computation[14], shows that some of these equilibria[15] can coexist in the sense that for a set of parameters, both types of equilibria occur:

Proposition 10. *There are sets of parameters, neither open nor closed, for which the following types of equilibria coexist (i.e. both can occur):*
1) PNT-HL(a_L) and PNT-LH(a_L).
2) PNT-HH(a_H) and PNT-LL(a_H).
3 FNT-H and FNT-L.
4) FNT-H and PNT-HH(a_H).
5) SEP and PNT-HH(a_H).

Typically, the question of coexistence of equilibria does not come up, since all of them always coexist (for example, in the Cho-Kreps beer-quiche game or Spencian signaling); they are, however, important in this setting since we will eventually apply refinements to select among these equilibria. If one views a refinement as simply a condition that a particular equilibrium may satisfy or not, the question of coexistence is irrelevant. If one views a refinement as a prediction of which of several equilibria is more plausible, one can conceivably say that if they do not coexist, one does not need a refinement to choose among equilibria, since the conditions for existence of an equilibrium will function as a kind of refinement. In either case, we show that the relevant equilibria do, in fact, coexist, so that a refinement has bite.

3.6 Discussion and Refinements

There are a number of notable differences between this simple model and the models presented by Hedlund (2016), Perez-Richet (2014) and Degan and Li (2015); one is the types of equilibria they admit. In Perez-Richet (2014)'s model separating equilibria are only possible when there exists a fully revealing experiment; otherwise all equilibria are pooling. In Hedlund (2016)'s model equilibria[16] are either pooling on the fully revealing experiment or fully separating where all types choose different experiments in equilibrium; furthermore the pooling and separating equilibria do not coexist. In the model discussed here nontrivial separating (in contrast to Perez-Richet (2014)) and equilibria where the pooling is on the less informative signal, as well as the striking feature

[13] It can be checked directly that there are no "perverse" equilibria where the receiver "inverts" the signal (that would never be optimal) or another separating equilibrium where the high type pretends to be the low type and vice versa.

[14] Using, for example, a computer algebra system such as Mathematica.

[15] There are other results on (non-)coexistence of various types of equilibria; we list only the ones that are relevant.

[16] He focuses on equilibria that also satisfy a refinement - criterion D1. In the present model this refinement does not make any predictions beyond those of PBE with tie-breaking.

of coexisting pooling and separating equilibria (in contrast to Hedlund (2016)) are possible. If, in addition, we dispense with the tie-breaking rule that is part of the present model, another, hybrid, type of equilibrium is possible, one where the type of sender randomizes, while the other plays a pure strategy. This type of equilibrium is not possible in either of the two alternative models. Degan and Li (2015) work in a setting that is similar to Perez-Richet (2014)'s, but posit type-independent costly signals; their results on the types of possible equilibria are analogous - in particular, there exists a unique separating equilibrium (which does not survive a refinement - D1 - which we also define shortly) in their model, and a number of pooling equilibria (which may or may not survive D1).

Previous work has also characterized equilibria of various models; in addition, owing to the fact that typically there are a large number of equilibria, various refinements have been brought to bear on the results, in order to obtain sharper predictions[17]. The most common refinement is criterion D1; we now give a suitably modified variant of its definition:

Definition 4 (Criterion D1). *Fix an equilibrium $\{p^*, q^*, \mu^*, \beta^*\}$, and let $u_S^*(\theta)$ the the equilibrium utility of each type of sender. For out-of-equilibrium pairs (Π', μ), let*
$D^0(\Pi', \theta) \triangleq \{\mu \in [\mathbb{P}(\omega_H|\theta_L), \mathbb{P}(\omega_H|\theta_H)] | u^*(\theta) = \hat{v}(\Pi, \mu^*, \theta) \leq \hat{v}(\Pi', \mu, \theta)]\}$, *and* $D(\Pi', \theta) \triangleq \{\mu \in [\mathbb{P}(\omega_H|\theta_L), \mathbb{P}(\omega_H|\theta_H)] | u^*(\theta) = \hat{v}(\Pi, \mu^*, \theta) < \hat{v}(\Pi', \mu, \theta)]\}$. *A PBE is said to survive criterion D1 if there is no θ' s.t.*

$$\{D(\Pi', \theta) \cup D^0(\Pi', \theta)\} \subsetneq D(\Pi', \theta') \tag{10}$$

Typically in signaling models this criterion is defined somewhat differently - in terms of receiver best responses, rather than beliefs; it is without loss in this setting to use this definition (see also Hedlund (2016)). In addition, it is usually defined using beliefs of the receiver about the type of the sender (here, ν), rather than the state of the world (μ) - this is due to the fact that in most other models, these are one and the same, while here they are distinct, and what matters for the payoff is the state of the world, hence the definition must be given in terms of that.

It can be checked by direct computation that all of the equilibria described above survive criterion D1, and thus, it does not help refine predictions beyond those of PBE with tie-breaking[18]. This is due to the fact that for all equilibria and deviations, criterion D1 requires a *strict* inclusion of the D sets, as emphasized in equation 9, while in this game the relevant D sets are, in fact, identical for both types. Similarly, other related refinements such as the intuitive criterion[19] and other

[17]Typically in cheap-talk games refinements based on stability have no bite since messages are costless. The standard argument for why that is true goes as follows: suppose that there is an equilibrium where a message, say m' is not sent, and another message, m, is sent. Then we can construct another equilibrium with the same outcome where the sender randomizes between m and m' and the beliefs of the receiver upon observing m' are the same as his beliefs upon observing m in the original equilibrium. Here this is not true - although all experiments are costless, they generate different signals with different probabilities. For the sender to be mixing she must be indifferent between both experiments, but given the different probabilities that is impossible, and therefore we cannot support all equilibria by mixing. Thus refinements based on stability and restricting beliefs "regain" their bite in this setting.

[18]Intuitively, D1 does not help due to the following: consider an equilibrium (and associated utility levels), and a deviation. The set of receiver beliefs that make one or both types better off is the set of beliefs for which the receiver takes the high action "more often" than in the reference equilibrium. But the set of these beliefs is *identical* for both types, since the receiver's utility only depends on the state of the world, and not on the type of the receiver.

[19]The reason this refinement does not work is that for the right range of beliefs both types benefit. Note also that

refinements based on strategic stability Kohlberg and Mertens (1986). Moreover, Other standard refinements for signaling games such as perfect sequential equilibria (Grossman and Perry (1986)), neologism-proof equilibria (Farrell (1993))[20], or perfect (Selten (1975)) or proper (Myerson (1978)) equilibria, also do not narrow down predictions, for similar reasons.

Finally, another refinement concept - undefeated equilibria (Mailath et al. (1993)) - does help refine equilibria somewhat. That refinement is defined for sequential equilibria, and it can be checked that all wPBE in this game can be sequential equilibria. Undefeated equilibrium still does not go far enough, as we will discuss after applying another refinement below.

The other related models have features that circumvent the problem of nonrefinability - in Hedlund (2016), it is the fact that the receiver's action is in a compact set, that the receiver's action is strictly increasing in the final belief, and the fast that the sender's utility is strictly increasing in the receiver's action[21]; in Perez-Richet (2014) it is the fact that sender is perfectly informed and the fact that the receiver can use mixed strategies; in Degan and Li (2015) it is the fact that the action of the sender (the message) is continuous and related to the precision of the signal observed by the receiver. We will say more about the differences between the present setting and others below.

There is, however, another, novel, refinement that we can define. Take for example the PNT-LH(a_L) equilibrium; one may notice that while other refinement concepts do not work well, there is a curious feature in this equilibrium. It is this: while neither types benefit from a deviation to Π_H under the equilibrium beliefs, and both types benefit from the same deviation under other, non-equilibrium beliefs, it is the *high* type that benefits *relatively* more. This observation suggests a refinement idea - one may restrict out-of-equilibrium beliefs to be consistent not just with the types that benefit (such as the intuitive criterion, neologism-proof equilibria and others) or sets of beliefs (or responses) of the sender for which certain types benefit (such as stability-based refinements), but also with the *relative* benefits from a deviation[22]. It is also hoped that this refinement will prove useful in other applications where other refinements perform poorly.

This idea is also connected to the idea of trembles (Selten (1975)); namely that if one thinks of deviations from equilibrium as unintentional mistakes, this can be accommodated by the present refinement, but with an additional requirement - the player for whom the difference between the equilibrium utility and the "tremble utility" is greater should tremble more, and therefore, the beliefs of the receiver should that that into account. A similar reasoning (albeit in a different setting) is also present in the justification for quantal response equilibrium (QRE) of McKelvey and Palfrey (1995) where players may tremble to out-of-equilibrium actions with a frequency that is proportional in a precise sense to their equilibrium utility. These ideas are also what is behind the nomenclature - BPM stands for Belief-Payoff Monotonicity. We now turn to this refinement, and show that it does help narrow down the predictions to some degree. We give a definition that is suitable to the present environment, but it can be generalized in a straightforward way.

were this not true, we would be in the range of parameters where the separating equilibrium occurs - c.f. SEP.

[20]Both of these two refinements also fail since both types benefit from a deviation under the same set of beliefs.

[21]We discuss in detail the differences between Hedlund's model and ours below.

[22]We further explore the implications, properties and performance of this criterion in related contemporaneous work.

Definition 5 (Criterion BPM)**.** *Let $\{p^*, q^*, \mu^*, \beta^*\}$ be an equilibrium and let $u^*(\theta)$ be the equilibrium utility of type θ. Define, for a fixed θ and Π_i, $\overline{v}(\theta_i) \triangleq \max_{a,\mu} \hat{v}(\Pi_i, \theta_i, \mu)$ and $\underline{v}(\theta_i) \triangleq \min_{a,\mu} \hat{v}(\Pi_i, \theta_i, \mu)$. An equilibrium is said to fail criterion BPM if there is an experiment Π_i, not chosen with positive probability in that equilibrium and a type of sender, θ_j, such that:*

i) Let $\hat{\mu} \in \Delta(\Omega)$ be an arbitrary belief of the receiver and suppose that $\delta(\Pi, \mu, \hat{\theta}_i, e) \triangleq \frac{\hat{v}(\Pi,\theta_i,\hat{\mu}) - u^*(\theta_i)}{\overline{v}(\theta_i) - \underline{v}(\theta_i)} > 0$, for that belief.

ii) Denote by K be the set of types for which (i) is true. Let θ_i be the type for which the difference is greatest. If there is another type θ_j in K, for which $\delta(\Pi, \mu, \theta_i, e) > \delta(\Pi, \mu, \theta_j, e)$ then let $\mu(\theta_j|\Pi) < \epsilon\mu(\theta_i|\Pi)$, for some positive ϵ, with $\epsilon < \frac{1}{|K|}$. If there is another type θ_k such that $\delta(\Pi, \mu, \theta_j, e) > \delta(\Pi, \mu, \theta_k, e)$, then let $\mu(\theta_k|\Pi) < \epsilon\mu(\theta_j|\Pi)$, and so on.

iii) Beliefs are consistent: given the restrictions in (ii), the belief $\hat{\mu}$ is precisely the beliefs that makes (i) true.

We say that an equilibrium fails the BPM criterion if it fails the ϵ-BPM criterion for every admissible ϵ. In words, criterion BPM restricts out-of-equilibrum beliefs of the receiver in the following way: if there are beliefs about off-equilibrium path deviations, for which one type benefits more than another, then equilibrium beliefs must assign lexicographically larger probability to the deviation coming from the type that benefits the most. We also scale the differences in a way that makes the definition ordinal (see also de Groot-Ruiz et al. (2013)). Note also that the second part of the definition looks very much like a condition of increasing differences; this is indeed so and purposeful. In addition, one can note that for utility functions which do satisfy increasing differences, criterion BPM would generate meaningful and intuitive belief restrictions.

The definition given above is ordinal (i.e., for any sender's vNM utility function $u(x)$ the definition has the same meaning if $u(x)$ was replaced by $v(x) = a + bu(x)$, for any real number a and any positive real number b).

From now on we will refer to a PBE with tie-breaking that also survives criterion BPM as a *BPM equilibrium*. We have the following proposition:

Proposition 11. *The following classes of equilibria are BPM equilibria: SEP, FNT-H, PNT-HL(a_L), PNT-HH(a_H) and PNT-LL(a_H).*

In other words, this proposition applies to parts 1 and 3 of proposition 10, and makes a selection between the coexisting equilibria mentioned there. It should be noted that these equilibria are also ϵ-BPM equilibria, for all admissible ϵ, but we suppress this fact in the exposition that follows. Perhaps an instructive graph may boost intuition.

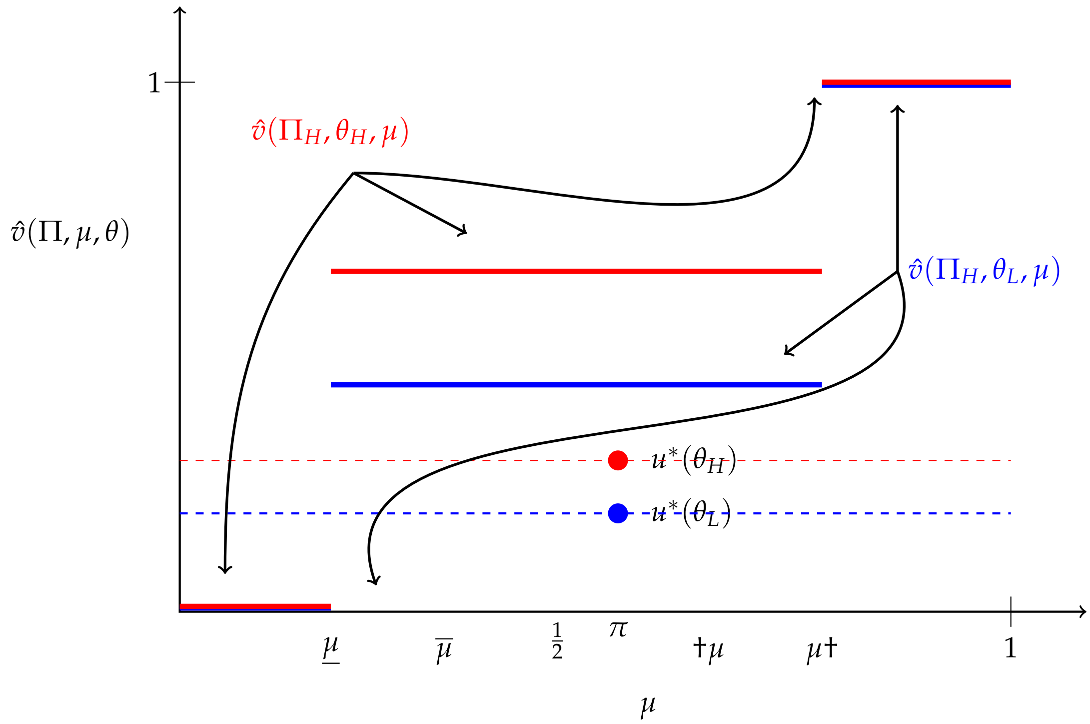

Figure 1: Illustration with pooling on Π_L, and the deviation to Π_H.

In the above figure the dots represent the on-path[23] utilities in the PNT-LH(a_L)[24] equilibrium for the high (red) type and the low (blue) type, and the dashed lines are there to make the comparisons of utilities from deviations easier; the equilibrium utility of deviating in that equilibrium is zero given the beliefs. The solid lines represent the expected utility of deviating to a more informative experiment as a function of the interim beliefs of the receiver; the differences between the solid and the dashed lines are computed in the proof above, for each μ. Clearly, for $\mu \in [0, \underline{\mu})$[25] both types get zero payoff from the deviation, since for those beliefs the receiver always takes the low action. Criterion BPM does not apply there since neither type benefits from such a deviation for those beliefs. The crucial region is $\mu \in [\underline{\mu}, \mu\dagger)$. It is here that criterion BPM operates efficiently - both types get positive payoff from the equilibrium and the deviation, but we have shown above that the high type benefits relatively more. And beliefs above $\mu\dagger$, again, cannot sustain a non-trivial equilibrium and hence we do not have to consider them since they lie outside the scope of admissible beliefs.

There is a small but important subtlety to be noticed - in *any* equilibrium (pooling or otherwise), $u_S^*(\theta_H) \geq u_S^*(\theta_L)$, because the private information of the sender (her type) forces the high type of the sender to have higher beliefs about the probability of higher signals, since $\mathbb{P}(\sigma_H|\theta_H) >$

[23]Here an throughout we use the terms "on-path" and "off-path" to mean objects (beliefs or actions) that are part of some equilibrium, but either occur on the path of play, or do not. We do not use terms like "out of equilibrium" since that could create confusion.

[24]A similar figure can constructed for the FNT-L equilibrium; it would be nearly identical except for the utility levels.

[25]Note that the right boundary is not included, since at that point the receiver would switch to taking the high action, by assumption.

$\mathbb{P}(\sigma_H|\theta_L)$. Nevertheless, given the restrictions on parameter discussed above, BPM does, in fact eliminate the equilibria where both types pool on the less informative experiments (with the exception of PNT-LL(a_H)); the reason it does not eliminate that equilibrium is because there, on the equilibrium path, the sender gets the highest possible utility she can get with probability one. Thus, no reasonable refinement could ever refine that outcome away, since the sender would never deviate from the equilibrium. As mentioned above, undefeated equilibrium does help to refine predictions, however, and in fact, makes a very similar selection, with the exception of FNT-L - some of the equilibria of that class are not refined away by undefeatedness.

Finitely many actions for the receiver and finitely many types for the sender can be accommodated easily in our setting; while we do not present explicit results to that end, it is straightforward to see that the same equilibria can exist in such an environment. We study an extension with an uncountable number experiments in the next section and show that analogous results continue to exist. Finally, to show that the results in our model do not depend on the *absence* of a fully revealing experiment, we explore this possibility. Interestingly, making Π_H be fully revealing in the present setting (i.e. setting $\rho_H = 1$) does not make much of a difference; the following (BPM) equilibria remain: $SEP, FNT-H, PNT-LL(a_L)$ and $PNT-HL(a_L)$.

3.7 Differences with the model of Hedlund: modeling assumptions and results.

As mentioned above, the model of Hedlund (2016) is rather close to the one discussed here; yet the predictions are sufficiently distinct. We now turn to a more detailed discussion of the differences (and similarities) between the models, as well as the implications of those differences for equilibria.

The most notable difference is that our model can support both pooling and separating equilibria, and even in BPM equilibria we can get pooling on the less informative experiment[26]. In addition, number of features of the equilibria in Hedlund (2016)'s model fail here; notably, the fact that in equilibrium the senders choose more informative experiments than they would have under symmetric information, as well as the fact that the payoff for senders is the same across all equilibria.

Finitely many actions for the receiver and finitely many types for the sender can be accommodated easily in our setting; while we do not present explicit results to that end, it is straightforward to see that the same equilibria can exist in such an environment. We study an extension with an uncountable number experiments in the next section and show that analogous results continue to exist. To show that the results in our model do not depend on the *absence* of a fully revealing experiment, we explore this possibility. Interestingly, making Π_H be fully revealing in the present setting (i.e. setting $\rho_H = 1$) does not make much of a difference; the following (BPM) equilibria remain: $SEP, FNT-H, PNT-LL(a_L)$ and $PNT-HL(a_L)$.

The assumptions that are responsible for these differences can be divided into two classes - assumptions about the actions available to the sender (i.e. the set of experiments), and assump-

[26]Recall that in Degan and Li (2015)'s model the D1 equilibria are also pooling.

tions about the utilities of the players as well as the actions available to the receiver. Changing the assumptions in either class will result in equilibria that are qualitatively closer to the equilibria of this model (notably, producing nontrivial pooling equilibria).

Consider first the assumptions regarding the set of available experiments. First of all, if the fully revealing experiment is not available in Hedlund (2016)'s model, the same results may not hold[27]; it should be noted that Perez-Richet (2014) also finds that absent a fully revealing experiment, there exist many PBEs, just like in the model we study. Another assumption is that all possible experiments are available to the sender, or equivalently, she can freely design them. This is crucial since some of the results rely on such a constructed experiment. However, suppose that we take Hedlund (2016)'s model and remove all experiments except for two - a fully revealing one, and an arbitrary other one. Then, if the common prior that the state is high is sufficiently close to 1, it will be an equilibrium for both types of sender to pool on the non-fully revealing experiment; moreover, this equilibrium will survive criterion D1, since both D^0 and D sets are empty. Thus, dropping the assumptions about the set of available experiments results in equilibria that are similar to the equilibria studied here.

Consider now the second class of assumptions. Among other differences between these models there are three key ones: *i*) a connected action space for the receiver, *ii*) the fact that the sender's utility is strictly increasing in the action of the receiver and *iii*) the fact that the receiver's best response is strictly increasing in the final belief. All three of these assumptions are not satisfied in the present setting. It is this combination of assumptions taken together that is responsible for the differences in results and predictions between the two models. We now show by examples that dropping any one of these four assumptions (but keeping the other three), and thus introducing some "coarseness" into the setting, would change the results of Hedlund (2016) significantly, elegant though they may be, and bring them closer to the results in this model.

One can also drop the assumption of a connected action set for the receiver: for convenience suppose that there are two types of sender, any finite number of available actions for the receiver and all other assumptions are the same as in Hedlund (2016). In this case the finite number of actions forces the possible utilities of the sender and receiver to also take on a finite number of values (and in addition, the receiver's optimal action can no longer be strictly increasing in his final belief, which is a key element in Hedlund (2016)) - therefore this effectively becomes analogous to the model studied in the present work, with all of the resulting conclusions.

Similarly, keeping a connected action space, and making $a^R(\beta)$ (the optimal action of the receiver as a function of his final belief) constant over some regions[28], or keeping $a^R(\beta)$ strictly increasing but making the sender's utility constant over some regions of the receiver's actions makes Hedlund (2016)'s results break down.

[27]It is not clear whether they do or do not but Hedlund's characterization would not apply.

[28]If this function is decreasing over some regions the model changes significantly, since then the preferences of the receiver are no longer about matching the state as closely as possible; we do not consider this case.

3.8 Welfare and Comparative Statics

We now turn to the question of welfare. For the receiver[29], the expected utility is the same across the FNT-H and PNT-HL(a_L) equilibria, and equal to $2\rho_H - 1$, which is positive by assumption. His utility from the equilibria FNT-L and PNT-LH(a_L) is strictly lower than that and equal to $2\rho_L - 1$. His utility from PNT-HH(a_H) and PNT-LL(a_H) is $2\pi - 1$. His utility from SEP is $(\rho_H - \rho_L)(3\pi\xi - 2\pi - 2\xi) + 2\rho_H - 1$; this can be positive or negative even in the range of relevant parameters. Thus among the pooling equilibria the receiver prefers the more informative one, and how he ranks the separating one is ambiguous. An interesting comparison is between the receiver's payoff in these equilibria and his payoff in the absence of any persuasion - that is, what the receiver would do based just on the prior. Clearly, if the prior is $\pi \geq \frac{1}{2}$ the receiver should take the high action, yielding a payoff of $2\pi - 1$ and if $\pi < \frac{1}{2}$, the receiver should choose the low action, and obtain $1 - 2\pi$ in expectation. One can definitely say in this case that if $\pi \geq \frac{1}{2}$ (and so, ex ante, the interests of the receiver and the sender are aligned), and the rest of the parameters are such that any type of pooling equilibrium obtains, the receiver strictly prefers the outcome under persuasion over that under no persuasion. This is a rather interesting result, showing that even if the sender always prefers one of the outcomes, the receiver may still prefer to be persuaded. Other utility comparisons are, again, ambiguous.

As for the sender, we can say that in any equilibrium, the expected utility of the high type is always weakly greater than that of the low type. Clearly the payoff for both types from PNT-HH(a_H) and PNT-LL(a_H) is equal to unity. The high type of sender obtains the same expected payoff from FNT-H, PNT-HL(a_L) and SEP; that payoff is equal to $\frac{\rho_H\pi\xi+(1-\rho_H)(1-\pi)(1-\xi)}{\pi\xi+(1-\xi)(1-\pi)}$. Her expected payoff from FNT-L and PNT-LH(a_L) is equal to $\frac{\rho_L\pi\xi+(1-\rho_L)(1-\pi)(1-\xi)}{\pi\xi+(1-\xi)(1-\pi)}$. As for the low type, her payoff from SEP, FNT-H, and PNT-HL(a_L) is $\frac{\rho_H\pi(1-\xi)+\xi(1-\rho_H)(1-\pi)}{\pi(1-\xi)+\xi(1-\pi)}$, and that FNT-L and PNT-LH(a_L) is: $\frac{\rho_L\pi(1-\xi)+\xi(1-\rho_L)(1-\pi)}{\pi(1-\xi)+\xi(1-\pi)}$. Comparing these expected payoffs is more difficult, since they involve all four parameters and different equilibria occur under different parameters; thus, it is not possible to say in general, which type of equilibrium each type prefers. However, when equilibria do coexist, the utility of FNT-H is higher than that of FNT-L for both types, and the same is true of PNT-HL(a_L) and PNT-LH(a_L). Thus, when it does make nontrivial selections, BPM picks out equilibria that are preferred by both the sender and the receiver. While BPM does not make a selection among PNT-HH(a_H) and PNT-LL(a_H), the sender clearly gets her first best in these equilibria. When these equilibria do coexist, the following figure summarizes the preferences of both types of the sender between them:

$$\left\{\begin{matrix} FNT-L \\ PNY-LH(a_L) \end{matrix}\right\} \preceq_{Sender} \left\{\begin{matrix} FNT-H \\ SEP \\ PNT-HL(a_L) \end{matrix}\right\} \preceq_{Sender} \left\{\begin{matrix} PNT-HH(a_H) \\ PNT-LL(a_H) \end{matrix}\right\}$$

It should be noted that the set of BPM equilibria is exactly the five equilibria denoted in the central and the right columns in the figure above. Notably, this is quite starkly different to the

[29] Note that for the specific utility function posited for the receiver, the expected utility of the receiver is also numerically equivalent to the probability of making the correct decision.

results of Hedlund, who shows that in a model where a perfectly revealing experiment is available the welfare of the sender is the same across all equilibria that survive a refinement.

3.9 Private information and persuasion

A natural question that one may ask is whether the sender benefits from private information in this setting - that is, whether the sender would ex-ante prefer to be informed or not. Without private information this model is identical to the model of KG, except for the available experiments. Without private information it also doesn't make sense to speak of the "type" of sender in this situation; therefore, without observing a private signal the sender would simply choose the more informative experiment, if the common prior π is above one half, and less informative experiment otherwise. The expected payoff for the sender would be equal to $\rho_H \pi + (1 - \rho_H)(1 - \pi)$, which is in between that of the high type and the low type. Thus we can conclude that the sender sometimes benefits from private information. This is in in line with Alonso and Camara (2016) who show that if a fully revealing experiment is available, the sender does not benefit from private information. In addition to lacking a fully revealing experiment, in this setting the private information of the sender is also not "redundant" in the sense that Alonso and Camara make precise in their work; this feature also allows an informed sender to be better or worse off. We also note that here the sender does not benefit from persuasion[30] (and in fact does strictly worse), if the receiver is ex-ante willing to take the high action (i.e. if $\pi \geq \frac{1}{2}$), and does strictly better otherwise. This observation has an analogue in KG - there, also, the sender benefits if the receiver is be willing ex-ante take an action that is inferior from the point of view of the sender.

3.10 Policy Implications

[COMPARISON WITH PLEADING THE FIFTH] Instructing the jurors on a grand jury to reason in a particular way - more specifically, the reasoning implied by the refinement - would provide incentives for prosecutors to produce as much evidence as possible. The fifth amendment was motivated by another concern - that of protecting the rights of citizens (but perhaps at a cost - that of loss of information). This policy prescription is different - it is concerned with eliciting information, and does not take into account the interests of the prosecutor (which are taken to be known[31])

3.11 Going Further: More Available Experiments

Armed with the setup and intuition from the preceding discussion, we can go somewhat further and dispense with arbitrarily restricting the set of available experiments to just two.

[30] In the sense of KG - that is, if the value function of the sender evaluated at the prior is greater than the expected payoff at the prior in the absence of any persuasion.

[31] In addition, the outcome of a trial may in general involve the defendant suffering personally; the prosecutor is under no such threat.

Suppose instead that a finite set of experiments was available, with the elements of that set still ranked according to the "more informative than" criterion (defined below). From the point of view of qualitative analysis, it is immaterial exactly how many experiments there are, as long as there are a finite number of them (and at least two) - the basic results about existence of a separating equilibrium and several types of pooling equilibria (one for each available experiment), along with the corresponding beliefs and parameter restrictions go through with the obvious adjustments. We do not present explicit results to that end.

Instead, consider now an uncountable set of experiments $\mathbf{\Pi}$ and endow it with the sup norm; suppose it is a closed and compact (in the natural topology associated with the sup norm) set, still ranked. More precisely, consider the set of 2×2 symmetric matrices that are parametrized by a single number - the probability of a correct signal in a state, denoted by ρ_i. Say that Π_i a generic experiment, letting $i \in I$ be some index set, and define a "more informative than"[32] order on the set of experiments as follows: if $i' \neq i$, $\Pi_{i'} \succ \Pi_i$ if and only if $1 > \rho_{i'} > \rho_i > \frac{1}{2}$. Denote by $\rho_a \triangleq \min_\rho \Pi$ and $\rho_b \triangleq \max_\rho \Pi$, so that $I = [a, b] \subset \mathbb{R}$ and let Π_A and Π_B be the corresponding experiments. Also, modify notation from the previous section slightly as follows: let $\hat{p}(\theta) \in \Pi$ and $p(\theta) \in \Delta(\Pi)$. Note that Π is convex (so that the existence result from the previous section applies).

Surprisingly, there are still only two classes of FNT pooling equilibria, one where pooling is on the most informative experiment and one where it is on the least informative one. This is due to the fact that the conditions for each type of sender that ensure no deviation from a particular Π_i upward (toward a more informative experiment) and downward (toward a less informative one) are incompatible (within the class of FNT equilibria), and thus, no equilibrium where the pooling is on Π_i s.t. $a < i < b$ exists.

Proposition 12. *There is a continuum of fully nontrivial pooling equilibria where $\hat{p}(\theta_H) = \hat{p}(\theta_L) = \Pi_b$. These equilibria exist as long as $\pi + \xi \geq 1, \pi \geq \xi, \tilde{\pi}\tilde{\rho_b} \geq 1, \rho_b > \pi, \tilde{\pi}\tilde{\rho_L} \geq \tilde{\tilde{\xi}}, \tilde{\rho_i}\tilde{\tilde{\xi}} > \tilde{\pi}, \forall i \in I \setminus b$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_i) \in [\mathbb{P}(\omega_H|\theta_L), \rho_a)$ for $i \neq b$. Denote this kind of equilibria by "FNT-b".*

Proposition 13. *There is a continuum of fully nontrivial pooling equilibria where $\hat{p}(\theta_H) = \hat{p}(\theta_L) = \Pi_a$. These equilibria exist as long as $\pi + \xi \leq 1, \pi \leq \xi, \tilde{\pi}\tilde{\rho_i}\tilde{\tilde{\xi}} \geq 1, \forall i \in I \setminus a, \rho_a > \pi, \tilde{\rho_a} > \tilde{\tilde{\xi}}\tilde{\pi}, \tilde{\rho_a}\tilde{\pi} \geq 1$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_i) \in [\mathbb{P}(\omega_H|\theta_L), \rho_b)$, for $i \neq a$. Denote this kind of equilibria by "FNT-a".*

There is also a unique separating equilibrium, which is analogous to the one constructed above.

Proposition 14. *There is a unique separating equilibrium where $\hat{p}(\theta_H) = \Pi_b, \hat{p}(\theta_L) = \Pi_a$. This equilibrium exists as long as $\pi \leq \xi, \pi + \xi > 1, \tilde{\pi}\tilde{\rho_b}\tilde{\tilde{\xi}} > 1, \tilde{\rho_b} > \tilde{\pi}\tilde{\tilde{\xi}}, \tilde{\pi}\tilde{\rho_a} > \tilde{\xi}, \tilde{\rho_a}\tilde{\xi} > \tilde{\pi}$. Denote this equilibrium by "SEP2".*

[32] This order is coarser (i.e. a subset of) both the "more precise than" order used by Hedlund, as well as Blackwell's standard order.

The reason that this is the only separating equilibrium is this. Suppose, to the contrary that there was another separating equilibrium, one where at least one type chose $\hat{p}(\theta) = \Pi_i$, for $\Pi_i \notin \{\Pi_a, \Pi_b\}$. Since the equilibrium is separating, that type would also reveal itself by it's choice, and thus $\mu(\Pi_i) = \mathbb{P}(\omega_H|\theta)$. The choice of that type of sender would then be

$$\max_{\Pi_i} \hat{v}(\Pi_i, \theta, \mathbb{P}(\omega_H|\theta)) \tag{11}$$

or, equivalently, given the structure of available experiments,

$$\begin{aligned} &\max_{\rho_i} \rho_i \left[\mathbb{P}(\omega_H|\theta)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta)\mathbb{1}_{|\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}}\right] + \\ &+(1-\rho_i)\left[\mathbb{P}(\omega_H|\theta)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}}\right] \end{aligned} \tag{12}$$

with $\mu = \mu(\Pi_i) = \mathbb{P}(\omega_H|\theta)$; the maximand is linear in ρ_i, and thus the solution is at one of the boundaries of the feasible set, and thus, for an equilibrium to be separating, each type must choose one of the "extreme" experiments[33]. Clearly, in a separating equilibrium they cannot choose the same one and it is not incentive compatible for the high type of sender to choose a very uninformative experiment, thus we arrive at the conclusion in the proposition.

There are two kinds of PNT equilibria, with continua of equilibria in each.

Proposition 15. *There is a continuum of p-nontrivial pooling equilibria where $\hat{p}(\theta_H) = \hat{p}(\theta_L) = \Pi_i, a^*(\Pi_i, \sigma) = a_H$, for $\sigma = \sigma_H, \sigma_L$ and $a^*(\Pi_j, \sigma_H) = a_H, a^*(\Pi_j, \sigma_L) = a_L$, for $i \neq j$. These equilibria exist as long as $\tilde{\rho}_j\tilde{\pi} \geq \tilde{\xi}, \rho_i \geq \pi, \tilde{\pi} < \tilde{\xi}\tilde{\rho}_j$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_j) \in \left[\mathbb{P}(\omega_H|\theta_j), \rho_L\right)$. Denote this kind of equilibria by "PNT-ii(a_H)".*

Proposition 16. *There is a continuum of p-nontrivial pooling equilibria where $\hat{p}(\theta_H) = \hat{p}(\theta_L) = \Pi_i, a^*(\Pi_i, \sigma_H) = a_H, a^*(\Pi_i, \sigma_L) = a_L$ and $a^*(\Pi_j, \sigma) = a_L$, for $\sigma = \sigma_H, \sigma_L, i \neq j$. These equilibria exist as long as $\rho_j > \pi$, $\rho_j + \pi \geq 1$ and $\tilde{\rho}_i\tilde{\pi} < \tilde{\xi}$. The only difference between these equilibria are the beliefs that the receiver holds off-path; namely, $\mu(\Pi_H) \in [\mathbb{P}(\omega_H|\theta_L), 1-\rho_i)$. Denote this kind of equilibrium by "PNT-ij(a_L)".*

Just like before, we have the following proposition:

Proposition 17. *There exist sets of parameters $\{\pi, \xi, \rho_a, \rho_b\}$ such that the following types of equilibria coexist:*
1) FNT-a and FNT-b.
2) There is a set $\tilde{I} \subseteq I$ such that for $i, i' \in \tilde{I}$, PNT-ii(a_H) and PNT-i'i(a_H) coexist.
3) There is a set $\tilde{I} \subseteq I$ such that for $i, i' \in \tilde{I}$, PNT-ij(a_H) and PNT-i'j(a_H) coexist.

And finally, analogously to the simpler model, we have the following result:

Proposition 18. *The following are BPM equilibria: SEP2, FNT-b, and for all $i \in I$, PNT-bi(a_L) and PNT-ii(a_H).*

[33] An elementary example of a "bang-bang" solution.

The argument for eliminating FNT-a, and PNT-ij(a_L) for $i \neq b$ is analogous to the argument given above for two experiments, and therefore omitted.

We end this section by noting simply that the results for two experiments extend to an uncountable set of experiments. Similar results can be obtained for the welfare of both the sender and the receiver.

4 A General Model: Non-dichotomous States.

There are a number of ways in which this basic model can be generalized; we present the one that is not typically pursued - a model with more than two states of the world.

Previous work on this problem was focused on a special case - the model presented earlier, as well as the models of Hedlund (2016), Degan and Li (2015) and Perez-Richet (2014) all focus on a binary state space - an assumption that is restrictive in the sense that the monotone likelihood ratio property and the single-crossing condition are "for free" in the sense that one can always put an order on the relevant set, perhaps with some renaming/relabeling of actions or signals, such that these properties hold. It would be interesting to consider more than two states - an extension to which we now turn. While we will not explicitly characterize the equilibria in detail as in section 2, we will show that criterion BPM operates in a similar way in such a setting.

4.1 General model.

Let $N \geq 2$ and $\mathcal{I}$ be an index set with N elements. Let $\Omega = \{\omega_i\}_{i \in \mathcal{I}}$, the set of states of the world, be the set of natural numbers less than or equal to N: $\Omega = \{1, 2, ..., N-1, N\}$.

Let $\Theta = \{\theta_1, ..., \theta_N\}$ be the set of types of receiver, let $\Sigma = \{\sigma_1, ..., \sigma_N\}$ be the set of signals, and let $A = \{a_1, ..., a_N\}$ be the set of actions for the receiver. We also identify Θ, Σ and A with the set of positive integers less than or equal to N, but for notational clarity will refer to elements of these sets using the corresponding nomenclature.

Let $\pi(\omega) \in \Delta(\Omega)$ be the common prior belief (probability mass function) about the true state, and denote by $F_\pi(\omega)$ the corresponding cumulative distribution function. The timing of the game is the same as in the simplified version. The sender receives a private signal according to a commonly known distribution $\xi(\theta|\omega)$; suppose for simplicity that $\forall \theta, \omega, \xi(\theta|\omega) > 0$. Upon seeing the realization of the type, the sender updates her beliefs to $\beta_S(\omega|\theta) \in \Delta(\Omega)$ as usual, according to Bayes rule: $\beta_S(\omega|\hat{\theta}) = \frac{\pi(\omega)\xi(\hat{\theta}|\omega)}{\sum_\omega \pi(\omega)\xi(\hat{\theta}|\omega)}$, along with the cumulative distribution $B_S(\omega)$[34]. The sender then chooses an information structure, $\Pi \in \mathbf{\Pi}$ which is a subset of $N \times N$ matrices (suppose also that $\mathbf{\Pi}$ is closed in the sup norm) of the following form: for $\rho \in [\underline{\rho}, \overline{\rho}]$, with $\frac{1}{2} < \underline{\rho} < \overline{\rho} < 1$, let Π_ρ be the experiment with ρ on the diagonal, and $\frac{1-\rho}{N-1}$ elsewhere. In other words,

[34] Throughout, capital letters will denote distribution functions and lower-case letter will denote probability mass functions.

$$\Pi_\rho = \begin{array}{c} \\ \sigma_1 \\ \sigma_2 \\ \sigma_3 \\ \vdots \\ \sigma_N \end{array} \begin{array}{c} \begin{array}{ccccc} \omega_1 & \omega_2 & \omega_3 & \dots & \omega_N \end{array} \\ \begin{pmatrix} \rho & \frac{1-\rho}{N-1} & \frac{1-\rho}{N-1} & \dots & \frac{1-\rho}{N-1} \\ \frac{1-\rho}{N-1} & \rho & \frac{1-\rho}{N-1} & & \\ \frac{1-\rho}{N-1} & \frac{1-\rho}{N-1} & \rho & & \\ \vdots & & & \ddots & \\ \frac{1-\rho}{N-1} & & & & \rho \end{pmatrix} \end{array}$$

We say that Π_ρ is *more informative* than $\Pi_{\rho'}$ iff $\rho > \rho'$. For convenience, denote the maximal element in $\mathbf{\Pi}$ by Π^*. The reason for focusing on this very special structure for experiments is due to the fact that other possible orders (Blackwell informativeness (Blackwell (1951), Blackwell (1953)) or Lehmann accuracy (Lehmann (1988), Persico (2000)) are either too general (such as Blackwell informativeness) or rather unsuitable to provide meaningful results in this setting (Lehmann accuracy). Similar results can be obtained for those more general and common orders, but they require very strong and difficult to interpret assumptions elsewhere, such as the utility function of the sender. Given interim beliefs $\mu(\omega|\Pi) \in \Delta(\Omega)$, the receiver updates to his final beliefs using Bayes rule. More precisely, suppose that the experiment chosen by the sender is Π, the interim belief is μ and the observed signal is σ_i. Then the final belief is simply

$$\beta(\Pi, \sigma_i, \mu) = \left(\frac{\Pi(\sigma_i|\omega_1)\mu(\omega_1|\Pi)}{\sum_j \Pi(\sigma_i|\omega_j)\mu(\omega_j|\Pi)}, \dots, \frac{\Pi(\sigma_i|\omega_N)\mu(\omega_N|\Pi)}{\sum_j \Pi(\sigma_i|\omega_j)\mu(\omega_j|\Pi)} \right)' \tag{13}$$

where the "prime" mark denotes the transpose of a vector; similarly the receiver computes final beliefs given any other signal.

The sender has state independent preferences, with (vNM) utility given by $u^S(a) : A \to [0,1]$, strictly increasing in a with $u^S(a_1) = 0$ and $u^S(a_N) = 1$. The receiver has (vNM) utility given by $u^R(a, \omega) : A \times \Omega \to \mathbb{R}$ with $u^R(a_i, \omega_i) = 1, \forall i = 1, ..., N$; thus, the receiver always wants to match the correct state. The utility of "mistakes" is given by $u(a_i, \omega_j) = 1 - |j - i|k$ for some $k \in (0,1]$.

For example, if $N = 5$,

$$u^R(a,\omega) = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{array} \begin{array}{c} \begin{array}{ccccc} \omega_1 & \omega_2 & \omega_3 & \omega_4 & \omega_5 \end{array} \\ \begin{pmatrix} 1 & 1-k & 1-2k & 1-3k & 1-4k \\ 1-k & 1 & 1-k & 1-2k & 1-3k \\ 1-2k & 1-k & 1 & 1-k & 1-2k \\ 1-3k & 1-2k & 1-k & 1 & 1-k \\ 1-4k & 1-3k & 1-2k & 1-k & 1 \end{pmatrix} \end{array}$$

An illustrative special case has $N = 3$ and $k = 1$

$$u^R(a,\omega) = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

We can view, for a fixed $a \in A$, u^R as a random variable, having the distribution F_π, M or B, depending on what the information of the receiver is at that point[35].

A pure strategy for the sender is a function $\hat{p}(\theta) : \Theta \to \mathbf{\Pi}$, and a mixed strategy is a distribution $p(\theta) : \Theta \to \Delta(\mathbf{\Pi})$; for convenience we identify a degenerate mixed strategy and a pure strategy, and write $p(\theta) = \delta_\Pi$ in that case, where δ_x is the Dirac distribution over $\mathbf{\Pi}$ centered at x. A pure strategy for the receiver is $\hat{q}(\Pi,\sigma) : \mathbf{\Pi} \times \Sigma \to A$ and a mixed strategy is $q(\Pi,\sigma) : \mathbf{\Pi} \times \Sigma \to \Delta(A)$; and similarly, denote by $q(\Pi,\sigma) = \delta_a$ a degenerate mixed (i.e. pure) strategy of playing action a.

Let $i > j$, and suppose that the family ξ satisfies the MLRP. We can make the following immediate

Observation 1. *The family of posteriors of the sender, $\beta_S(\omega|\theta)$, are ranked according to the FOSD order (Milgrom (1981)). In other words, for $\omega_i > \omega_j$, and $\theta_i > \theta_j$,*

$$\frac{\xi(\theta_i|\omega_i)}{\xi(\theta_i|\omega_j)} \geq \frac{\xi(\theta_j|\omega_i)}{\xi(\theta_j|\omega_j)} \Rightarrow \; B_S(\omega|\theta_i) \succ_{FOSD} B_S(\omega|\theta_j) \tag{14}$$

In other words, a higher observed signal type for the sender is always "good news" in the sense of FOSD.

From now on we will focus only on pure strategies, for both sides of the game, to simplify the analysis; again, suppose that the receiver breaks any ties in favor of the higher action, so that the sender's expected utility function is upper-semi-continuous. This assumption is rather less than innocuous, since one might lose the existence of equilibrium, in addition to narrowing down the scope of possibilities. Nevertheless we are forced to make it to solve the game, as well as to extend the results clearly; from now on, write $p(\theta) = \Pi$, for some $\Pi \in \mathbf{\Pi}$, and $q(\Pi,\sigma) = a$, for $a \in A$. We can extend the definition of fully nontrivial, partially nontrivial and pooling equilibria in a straightforward way.

Suppose that the receiver holds final beliefs $\beta(\omega|\Pi,\sigma,\mu)$. The problem facing him at that point is

$$\max_{a\in A} \sum_j u^R(a,\omega_j)\beta(\omega_j|\Pi,\sigma,\mu) \tag{15}$$

which is clearly just maximizing the expected value of the random variable u^R by choice of a. Let $a^*(\Pi,\sigma,\mu)$ or, equivalently, $a^*(\beta)$[36] denote the solution. Suppose that in the case a tie, the receiver

[35]We write this having in the back of our minds a probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ with a finite number of outcomes and a state space $\{\mathbb{R}, \mathcal{B}(\mathbb{R})\}$ where $\mathcal{F}$ is just 2^Ω, the probability measure $\mathbb{P}$ may be π, μ, β_S, β, and $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.

[36]Hopefully the abuse of notation does not create confusion.

chooses the higher action; this assumption along with the specification of preferences yields the observation that the receiver's best response is always a pure strategy. The following lemma, the proof of which is the appendix, is not necessary for out analysis, but interesting in it's own right, given that the preferences of the receiver aren't just to take higher actions - they are to take the correct action:

Lemma 4.1. *The function $\beta \mapsto a^*(\beta)$ is weakly increasing in the following sense: if $B' \succ_{FOSD} B$, then either $a^*(\beta') \succ_A a^*(\beta)$ or $a^*(\beta') = a^*(\beta)$.*

We can similarly define a function that gives each type's expected payoff for a fixed interim belief μ as follows:

$$\hat{v}(\Pi, \theta_i, \mu) \triangleq \mathbb{E}_\omega \left(\mathbb{E}_\sigma (u^S(a^*(\beta(\Pi, \sigma, \mu))))|\theta_i \right) = \sum_k \beta_S(\omega_K|\theta_i) \sum_j u^S(a^*(\beta(\Pi, \sigma_j, \mu)))\Pi(\sigma_j|\omega_k) \tag{16}$$

Optimality requires that for each θ_i,

$$\hat{\Pi} \in \arg\max_{\Pi \in \mathbf{\Pi}} \hat{v}(\hat{\Pi}, \theta_i, \mu(\hat{\Pi})) \tag{17}$$

We can make several observations about $\hat{v}$. First, for a fixed Π, and i, if $M' \succ_{FOSD} M$, then $\hat{v}(\Pi, \theta_i, \mu') \geq \hat{v}(\Pi, \theta_i, \mu)$; this follows from Observation 1 and Lemma 4.1. In other words, ceteris paribus, a more optimistic interim belief is unequivocally beneficial for the sender. Second, for a fixed Π and μ, $i > l$, then $\hat{v}(\Pi, \theta_i, \mu) \geq \hat{v}(\Pi, \theta_l, \mu)$.

We have so far omitted a discussion of the role of the interim beliefs of the receiver about the type of sender: $\nu = \nu(\theta|\Pi) \in \Delta(\Theta)$. It plays the same role, specifying the equilibrium beliefs of the receiver, according to the strategies of the different types of sender.

The first basic observation that we can make is about existence of fully pooling[37] equilibria; while we make a strong assumption about π and ξ in doing so, this is just to give a sufficient condition that is both simple, and works across different other parameters:

Proposition 19. *Suppose that π and ξ are such that for all ρ, $\beta_S(\omega_1|\theta_1) \geq \frac{\rho(N-1)\beta(\omega_N|\theta_1)}{1-\rho}$. Then there exist fully pooling equilibria.*

Proof. As usual, we support such equilibria by extremely "pessimistic" beliefs. Suppose that $\hat{p}(\theta_i) = \Pi_\rho$ for all i, for some Π_ρ. Thus, on the equilibrium path $\mu(\Pi_\rho) = \pi$ and suppose that in case of a deviation the receiver believes that it came from the lowest type: $\mu(\Pi_{\rho''}) = \beta(\omega|\theta_1)$, $\rho'' \neq \rho'$. Then, given the restriction in the statement of the proposition, the receiver will find it optimal to take the lowest action, a_1, regardless of the signal. For all types of the sender this entails a utility of zero, and thus, this deviation will not be profitable. □

While we don't know what the function $\sigma \mapsto a^*(\sigma)$ looks like in general, without still further assumptions, we can make the following useful definition:

[37] We focus on *fully* pooling equilibria, namely those where all types of sender use the same pure strategy. There may exist others, with some pooling and some separation, but for the purposes of applying criterion BPM, there is no difference whether an equilibrium involves separation by some types or not.

Definition 6 ("Kind" of an equilibrium)**.** *Let e' and e'' be two equilibria. We say that these equilibria are of the same kind if in each equilibrium, on and off the equilibrium path, the mapping $\sigma \mapsto a^*(\sigma)$ between the realized signal in experiments that are chosen with any probability (including zero) is the same.*

This definition generalizes the nomenclature for the kinds of equilibria encountered in the simple model (PNT vs FNT, etc) and adapts it to a case with many actions and many states. We also assume that different equilibria of the same kind coexist. Since the simplest model discussed in the beginning is a special case of this one, we know that equilibria can, in fact coexist.

Instead of fully characterizing all equilibria, and then applying a refinement, we now focus just on pooling equilibria, and show that BPM operates in a similar and attractive way in a setting with a non-dichotomous state. A full characterization is available, but is not any more enlightening than in the case with two states. To that end, suppose that ξ and $\mathbf{\Pi}$ are such that there is a continuum of FNT equilibria[38].

We state two versions of the following result; one is for fully pooling, fully nontrivial equilibria, and the second for fully pooling equilibria of the same kind. The former is much easier to state and prove, and doesn't require additional notation, but it is subsumed by the latter.

Theorem 4.2. *Suppose that $\{p, q, \mu, \beta\}$ is a fully pooling, fully nontrivial equilibrium in which $\forall\theta$, $\tilde{p}(\theta) = \Pi_\rho$ and $\Pi_\rho \prec \Pi^*$. Then this equilibrium fails criterion BPM. Moreover, the unique equilibrium that survives criterion BPM among the class of fully pooling, fully nontrivial equilibria is one where $p(\theta) = \Pi^*$.*

Before we state the general version of this theorem, we need an additional definition.

Definition 7 ("Rank" of an action)**.** *Let e be a fully pooling equilibrium. The rank of an action, denoted by $n(a)$ is given by the following expression: $n(a) \triangleq card\{\sigma | a^*(\sigma) = a\}$ on the equilibrium path.*

In other words, the rank of an action is the number of signals that lead to that action on the equilibrium path. In particular, in a fully nontrivial equilibrium the rank of each action is equal to unity. We have the following immediate observation, the proof of which stems from comparing the definitions of kind and rank, and which we thus omit - if two equilibria are of the same kind, then all receiver actions in those equilibria have the same rank, but the converse is not necessarily true.

Theorem 4.3. *Suppose that e' and e'' are two fully pooling equilibria of the same kind, with pooling on $\Pi_{\rho'}$ and Π_ρ, respectively; suppose also $\rho' > \rho$. Suppose that the receiver takes at least two different actions on the equilibrium path and that the maximum rank of any action is bounded[39] above by $\frac{N}{3}$. The unique (among equilibria of the same kind) equilibrium that survives criterion BPM is the equilibrium where the pooling is on the most informative experiment, $\Pi_{\rho'}$.*

[38] It is possible to give explicit conditions that would guarantee this, but assuming those conditions would be equivalent to assuming this, and not elucidate anything in addition, so we are not explicit about them.

[39] We can give a weaker bound, and in fact, it will be apparent in the proof, but this is a convenient uniform, albeit stronger bound that also works.

The proofs of the theorems are in the appendix; it goes along the same lines as the two-state case - computing the relevant utilities. Note also that this definition generalizes the selection among PNT and FNT equilibria encountered in the simple model; there, too, criterion BPM was used to select among the different kinds of equilibria. Notably, however, in the simple model criterion BPM could not select between some equilibria simply because they did not coexist for the same parameters, and thus the question of selection among them was meaningless. While this can also happen in a more general setting for some specification of π, ξ and $\mathbf{\Pi}$, if different kinds of equilibria do coexist, we expect criterion BPM to operate in the same way and select the equilibria with the most revelation of information. A proof of this statement would rely on a particular specification, and lacking one, we do not give it.

We conclude this section by noting that the results of the model in this section are rather similar to the simpler model, as was expected. Not only does criterion BPM apply in a setting with more than two states, but it also operates in a manner that is analogous to that of the setting with a binary state.

5 Concluding Remarks

We present a relatively simple and straightforward model of communication between an imperfectly informed sender who is trying to persuade a receiver to take a certain action. The model differs somewhat from existing work, yet is tractable enough to derive similar (and in some cases, stronger) results. We work with a basic example using a particular specification of preferences and available information structures, that allows us to make reasonably strong predictions. We further refine the predictions using a novel yet intuitive refinement concept.

There are a number of directions in which this model can be extended in a fruitful way. For example, the sets of available experiments may vary with the state. This introduces an additional consideration for the receiver - if he doesn't see a certain signal, does that mean that the sender chose not to send it, or is it because it is not available? A similar restriction can apply to the types of sender; in the general model these restrictions would be manifested by conditions on ξ and $\mathbf{\Pi}$.

As a final note, and another way forward for future research, Hedlund (2016) shows that in his setting with $N \geq 2$ types, focusing on only two signals actually does involve some loss of generality; we appeal to the work of Taneva (2016) to argue that in general, one can restrict attention to "direct" experiments; however, it remains unclear if the restriction to symmetric experiments, and ones that are ranked by the "more precise than" criterion leads to any loss of generality.

Alonso and Camara (2016b) show that if a fully revealing information structure is available, then an uninformed sender (i.e. before, or without observing a private signal, in this paper, θ) can replicate any distribution of payoffs that can be achieved by an informed sender, and therefore, in a sense, private information is not useful in that setting. Their result does not apply to this model; this is to say that in realistic settings the sender will, in general, be able to manipulate the actions of the receiver based on what she knows.

Thus, while the assumption of the existence of a perfectly revealing experiment allows for characterization of equilibria, it also generates very specific results. More generally, it seems to be emerging from this and similar models that the mere presence or availability of a fully revealing experiment is one of the key features (among others, as discussed above) that drive results. In recent work on multi-sender persuasion an interestingly similar insight has emerged - the capability of one player to unilaterally mimic a particular distribution of signals (which can be thought of as an analogue to a fully revealing experiment in a single-sender framework) has become a key condition.

Appendix A: Proofs

Proof of Proposition 11. First, it is immediate that SEP is a BPM equilibrium, since there are no out-of-equilibrium beliefs to consider, and thus criterion BPM is trivially satisfied. The reason that PNT-LL(a_H) and PNT-HH(a_H) survive criterion BPM (note also that from proposition 10 we know that they coexist, so it is meaningful to talk about choosing between them) is that deviations from those equilibria do not yield a strictly higher payoff for either type. The computation that eliminates FNT-L and PNT-LH(a_L) goes as follows: Take any pooling equilibrium where both both types choose the experiment Π_L and the receiver takes different actions on the equilibrium path. In that equilibrium, $u^*(\theta_H) =$

$$\begin{aligned}\hat{v}(\Pi_L, \pi, \theta_H) = \rho_L \left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_L)\left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right]\end{aligned} \tag{18}$$

and $u^*(\theta_L) =$

$$\begin{aligned}\hat{v}(\Pi_L, \pi, \theta_L) = \rho_L \left[\mathbb{P}(\omega_H|\theta_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_L)\left[\mathbb{P}(\omega_H|\theta_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right]\end{aligned} \tag{19}$$

Fix a μ and consider the utility of deviating to Π_H for both types:

$$\begin{aligned}\hat{v}(\Pi_H, \mu, \theta_H) - u^*(\theta_H) = \rho_H \left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{|\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_H)\left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}}\right] - \\ -\rho_L\left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_L)\left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right] = \\ = (\mathbb{P}(\omega_H|\theta_H))\,[\rho_H\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} - \rho_L\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}} + (1-\rho_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}} - \\ -(1-\rho_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}}] + (\mathbb{P}(\omega_L|\theta_H))[\rho_H\mathbb{1}_{|\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}} - \rho_L\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}} + \\ +(1-\rho_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} - (1-\rho_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}] \geq 0, \forall\mu\in[0,1]\end{aligned} \tag{20}$$

Now let $\underline{\mu}$ solve $\frac{\rho_H\underline{\mu}}{\rho_H\underline{\mu}+(1-\rho_H)(1-\underline{\mu})} = \frac{1}{2}$, (i.e. $\underline{\mu} = 1-\rho_H$) and let $\bar{\mu}$ solve $\frac{\rho_L\overline{\mu}}{\rho_L\overline{\mu}+(1-\rho_L)(1-\overline{\mu})} = \frac{1}{2}$ (i.e. $\underline{\mu} = 1-\rho_L$) and note that since $\rho_H > \rho_L$, $\underline{\mu} < \overline{\mu}$. Also let $\dagger\mu$ solve $\frac{(1-\rho_L)\dagger\mu}{(1-\rho_L\dagger\mu+\rho_L(1-\dagger\mu))} = \frac{1}{2}$ (i.e. $\dagger\mu = \rho_L$) and $\mu\dagger = \frac{(1-\rho_H)\mu\dagger}{(1-\rho_H)\mu\dagger+\rho_H(1-\mu\dagger)} = \frac{1}{2}$ (i.e. $\mu\dagger = \rho_H$) and note that $\dagger\mu < \mu\dagger$. As before, we focus on nontrivial equilibria (so that we can disregard the terms that involve observing the low action). Now we can directly compute

$$
\begin{aligned}
\hat{v}(\Pi_H, \theta_H, \mu) - u^*(\theta_H) - (\hat{v}(\Pi_H, \theta_L, \mu) - u^*(\theta_L)) = \\
= [\mathbb{P}(\omega_H|\theta_H) - \mathbb{P}(\omega_H|\theta_L)] \left[\rho_H \mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} - \rho_L \mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right] + \\
+ [\mathbb{P}(\omega_L|\theta_H) - \mathbb{P}(\omega_L|\theta_L)] \left[(1-\rho_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} - (1-\rho_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right] = \\
= \begin{cases} u^*(\theta_L) - u^*(\theta_H) < 0, & \text{for } \mu \in [0, \underline{\mu}) \\ 2(\rho_H - \rho_L)(\mathbb{P}(\omega_H \theta_H) - \mathbb{P}(\omega_H|\theta_L))) > 0 & \text{for} \mu \in [\underline{\mu}, \dagger\mu) \\ 2\rho_L[\mathbb{P}(\omega_H|\theta_L) - \mathbb{P}(\omega_H|\theta_H)] + \mathbb{P}(\omega_H|\theta_H) - \mathbb{P}(\omega_H|\theta_L) < 0 & \text{for } \mu \in [\dagger\mu, 1] \end{cases}
\end{aligned}
\tag{21}
$$

Since the difference is negative for first of the three ranges exhibited above, criterion BPM does not apply there. For the second range of beliefs the difference is strictly positive, and hence, beliefs that support FNT-L or PNT-LH(a_L) are ruled out. As for the third range, the difference is negative, but beliefs there are such that they cannot be part of any kind of nontrivial equilibrium at all (cf. the upper bounds on off-path beliefs for equilibria in Propositions 4 through 9 and note that criterion BPM restricts beliefs off the equilibrium path) and we are done. □

Proof of Lemma 4.1. We first state the following common lemma (which is the discrete version of integration by parts) without proof:

Lemma 5.1. *(Abel's lemma)*
Let $\{a_i\}_{i=1}^n$ and $\{b_i\}_{i=1}^n$ be two sequences of real numbers. Let $A_i = \sum_{j=1}^i a_j$ and $B_i = \sum_{j=1}^i b_j$. Then

$$
\sum_i^n a_i b_i = \sum_i^{n-1} A_i(b_i - b_{i+1}) + A_n b_n \tag{22}
$$

Suppose that $B'(\omega) \succ_{FOSD} B(\omega)$ and fix take any a', a with $a' > a$. Consider the following difference:

$$
\begin{aligned}
\left[\sum_j u(a', \omega_j)\beta'(\omega_j) - \sum_j u(a, \omega_j)\beta'(\omega_j)\right] - \left[\sum_j u(a', \omega_j)\beta(\omega_j) - \sum_j u(a, \omega_j)\beta(\omega_j)\right] = \\
= \sum_j^{N-1} \left(B'(\omega_j) - B(\omega_j)\right) \left[u(a', \omega_j) - u(a, \omega_j) - u(a', \omega_{j+1}) + u(a, \omega_{j+1})\right]
\end{aligned}
\tag{23}
$$

where the equality is just applying Abel's lemma to appropriately defined variables, and the the fact that $B'(\omega_k) = \sum_{i=1}^k \beta'(\omega_i)$ and $B(\omega_k) = \sum_{i=1}^k \beta'(\omega_i)$ are discrete distribution functions. Given the utilities, it can then be checked by direct computation that the term is the square brackets weakly increasing in ω; this, combined with the fact that $\beta' \succ_{FOSD} \beta$ shows that the entire expression is nonnegative. In other words, that the function $f(a, \beta) \triangleq \mathbb{E}_\beta u(a, \omega)$ has increasing differences in (a, β). The fact that $a^*(\beta') \succ_A a^*(\beta)$ or $a^*(\beta') = a^*(\beta)$ for $\beta' \succ_{FOSD} \beta$ then follows by a standard argument. Namely, the choice set is totally ordered (a one-dimensional "chain", so that supermodularity trivially holds), the set of beliefs is a partially ordered set according to FOSD and f has increasing differences (and so also satisfies the single crossing condition). Thus, $a^*(\beta)$ is monotone nondecreasing in β (Milgrom and Shannon (1994)), and we are done. $\square$

Proof of Theorem 4.2. We can directly compute the required utilities, as follows. Given that the equilibrium is fully nontrivial (i.e. the function $\sigma \mapsto a_j(\sigma)$ is a bijection and $a^*(\sigma_j) = a_j, \forall j$)), the expected utilities of types i and $l, i > l$ for the sender are simply[40]

$$u^*(\theta_i) = \hat{v}(\Pi_\rho, \mu(\Pi_\rho), \theta_i) = \sum_k \beta_S(\omega_k|\theta_i) \sum_j u^S(a_j^*(\beta(\omega|\Pi_\rho, \sigma_j, \pi)))\Pi_\rho(\sigma_j|\omega_k) \tag{24}$$

and

$$u^*(\theta_l) = \hat{v}(\Pi_\rho, \mu(\Pi_\rho), \theta_l) = \sum_k \beta_S(\omega_k|\theta_l) \sum_j u^S(a_j^*(\beta(\omega|\Pi_\rho, \sigma_j, \pi)))\Pi_\rho(\sigma_j|\omega_k) \tag{25}$$

noting that $a_j^*(\beta(\omega|\Pi_\rho, \sigma_j, \pi)) = \sigma_j$, and using $\mu(\Pi_\rho) = \pi$ since the equilibrium is fully pooling). Now consider the expected utilities for a deviation to Π', using as $\mu = \mu(\Pi_{\rho'})$ any interim belief for which (a) the higher type of sender benefits from the deviation, and (b) is such that the equilibrium is still fully nontrivial. The "off-path" utilities are (we use the term "off path" in quotes, since it is off-path for the equilibrium which involves pooling on Π_ρ), but there is a coexisting equilibrium where this action is on-path):

$$\hat{v}(\Pi_{\rho'}, \theta_i, \mu) = \sum_k \beta_S(\omega_k|\theta_i) \sum_j u^S(a_j(\beta(\omega|\Pi_{\rho'}, \sigma_j, \mu)))\Pi_{\rho'}(\sigma_j|\omega_k) \tag{26}$$

and

$$\hat{v}(\Pi_{\rho'}, \theta_l, \mu) = \sum_k \beta_S(\omega_k|\theta_l) \sum_j u^S(a_j(\beta(\omega|\Pi_{\rho'}, \sigma_j, \mu)))\Pi_{\rho'}(\sigma_j|\omega_k) \tag{27}$$

Now compute the difference:

$$\begin{aligned} &\hat{v}(\Pi_{\rho'}, \theta_i, \mu) - u^*(\theta_i) - \left[\hat{v}(\Pi_{\rho'}, \theta_l, \mu) - u^*(\theta_l)\right] = \\ &= \sum_k (\beta_S(\omega_k|\theta_i) - \beta_S(\omega_k|\theta_l)) \sum_j u^S(a_j) \left[\Pi_{\rho'}(\sigma_j|\omega_k) - \Pi_\rho(\sigma_j|\omega_k)\right] = \\ &= \sum_k (\beta_S(\omega_k|\theta_i) - \beta_S(\omega_k|\theta_l)) \left((\rho' - \rho)u^S(a_k) + (\rho - \rho') \sum_{j \neq k} u^S(a_j))\right) \end{aligned} \tag{28}$$

the first equality relies on the fact that we are focusing on fully nontrivial equilibria and can thus dispense with keeping track of the disparate interim beliefs. Note the use the fact that the function $\sigma \mapsto a_j(\sigma)$ is monotonic, one-to-one and onto in collecting terms that involve utility.

Fix a state, say, ω_k. It can be checked by direct calculation that the expression $(\rho' - \rho)u^S(a_k) + (\rho - \rho') \sum_{j \neq k} u^S(a_j))$ is increasing in ω, for a fixed ω. Namely, take ω_k and ω_{k+1} and let $\phi(\rho, \rho', \omega_k) \triangleq$

[40]Note that by assumption the best the sender could obtain is one, and the worst is zero; this makes the calculation easier since we do not have to keep track of the normalization implicit in the definition of Criterion BPM.

$(\rho' - \rho)\left(u^S(a_k) - \sum_{j\neq k} u^S(a_j))\right)$ consider the difference $\phi(\rho,\rho',\omega_{k+1}) - \phi(\rho,\rho',\omega_k) = 2(\rho'-\rho)(u^S(a_{k+1}) - u^S(a_k))) > 0$ since $\rho' > \rho$ by supposition, and u^S is strictly increasing. Therefore, since $B_S(\omega|\theta_i) \succ_{FOSD} B_S(\omega|\theta_l)$, the entire expression is nonnegative by definition of first-order stochastic dominance. Since this same argument can be repeated until we arrive at $i = N$ and $\Pi_{\rho'} = \Pi^*$, we are done. $\square$

Proof of Theorem 4.3. We again compute the relevant utilities. In the baseline equilibrium the utilities are

$$u^*(\theta_i) = \hat{v}(\Pi_\rho, \mu, \theta_i) = \sum_k \beta_S(\omega_k|\theta_i) \sum_j u^S(a_j) \sum_{m\leq j} \Pi_\rho(\sigma_m|\omega_k)\mathbb{1}_{\{\sigma_m|a^*(\sigma_m)=a_j\}} \tag{29}$$

and

$$u^*(\theta_l) = \hat{v}(\Pi_\rho, \mu, \theta_l) = \sum_k \beta_S(\omega_k|\theta_l) \sum_j u^S(a_j) \sum_{m\leq j} \Pi_\rho(\sigma_m|\omega_k)\mathbb{1}_{\{\sigma_m|a^*(\sigma_m)=a_j\}} \tag{30}$$

and the utilities from the deviation are

$$\hat{v}(\Pi_{\rho'}, \mu, \theta_i) = \sum_k \beta_S(\omega_k|\theta_i) \sum_j u^S(a_j) \sum_{m\leq j} \Pi_{\rho'}(\sigma_m|\omega_k)\mathbb{1}_{\{\sigma_m|a^*(\sigma_m)=a_j\}} \tag{31}$$

and

$$\hat{v}(\Pi_{\rho'}, \mu, \theta_l) = \sum_k \beta_S(\omega_k|\theta_l) \sum_j u^S(a_j) \sum_{m\leq j} \Pi_{\rho'}(\sigma_m|\omega_k)\mathbb{1}_{\{\sigma_m|a^*(\sigma_m)=a_j\}} \tag{32}$$

Taking the difference in utilities between the different experiments for one type of sender yields

$$\begin{aligned} &\hat{v}(\Pi_{\rho'}, \mu, \theta_i) - u^*(\theta_i) = \\ &= \sum_k \beta_S(\omega_k|\theta_i) \sum_j (\rho'-\rho) u^S(a_k)\mathbb{1}_{\{\sigma_k|a^*(\sigma_k)=a_k\}}) + \left(\frac{\rho-\rho'}{N-1}\right)\left[(u^S(a_k)(n(a_k) - \mathbb{1}_{\{\sigma_k|a^*(\sigma_k)=a_k\}}) + n(a_j)u^S(a_j)\right] \end{aligned} \tag{33}$$

Now taking the difference between the utilities between different sender types

$$\begin{aligned} &\hat{v}(\Pi_{\rho'}, \mu, \theta_i) - u^*(\theta_i) - \hat{v}(\Pi_{\rho'}, \mu, \theta_l) + u^*(\theta_l) = \\ &= \sum_k (\beta_S(\omega_k|\theta_i) - \beta_S(\omega_k|\theta_l)) \Big[(\rho'-\rho))(u^S(a_k)(n(a_k) - \mathbb{1}_{\{\sigma_k|a^*(\sigma_k)=a_k\}} + \\ &\quad + \sum_j \left(\frac{\rho-\rho'}{N-1}\right)\left[(u^S(a_k)(n(a_k) - \mathbb{1}_{\{\sigma_k|a^*(\sigma_k)=a_k\}}) + n(a_j)u^S(a_j)\right]\Big] \end{aligned} \tag{34}$$

Now letting

$$\begin{aligned} &\hat{\phi}(\rho,\rho',\omega_k) \triangleq \\ &\triangleq \left[(\rho'-\rho))(u^S(a_k)(n(a_k) - \mathbb{1}_{\{\sigma_k|a^*(\sigma_k)=a_k\}}) + \sum_j \left(\frac{\rho-\rho'}{N-1}\right)\left[(u^S(a_k)(n(a_k) - \mathbb{1}_{\{\sigma_k|a^*(\sigma_k)=a_k\}}) + n(a_j)u^S(a_j)\right]\right] \end{aligned} \tag{35}$$

be the function that gives the expected utility of deviation as a function of the state and parameters, it can once again be checked directly that $\hat{\phi}(\rho, \rho', \omega_{k+t}) - \hat{\phi}(\rho, \rho', \omega_k) \geq 0$ for $t = 1, 2, ..., N-k$. There are six cases to consider (this is also where the condition $n(a) \leq \frac{N}{3}$ emerges from):

1. $n(a_k) = n(a_{k+t}) > 0$; in this case the expression $\hat{\phi}(\rho, \rho', \omega_{k+t}) - \hat{\phi}(\rho, \rho', \omega_k)$ is positive as long as $n(a_k) = n(a_{k+1}) \leq \frac{N}{3}$.

2. $n(a_k) = n(a_{k+t}) = 0$; in this case the expression $\hat{\phi}(\rho, \rho', \omega_{k+t}) - \hat{\phi}(\rho, \rho', \omega_k)$ is zero.

3. $n(a_k) > n(a_{k+t}) > 0$; in this case the expression $\hat{\phi}(\rho, \rho', \omega_{k+t}) - \hat{\phi}(\rho, \rho', \omega_k)$ is positive.

4. $n(a_{k+t}) > n(a_k) > 0$; in this case the expression $\hat{\phi}(\rho, \rho', \omega_{k+t}) - \hat{\phi}(\rho, \rho', \omega_k)$ is positive.

5. $n(a_k) > n(a_{k+t}) = 0$; in this case the expression $\hat{\phi}(\rho, \rho', \omega_{k+t}) - \hat{\phi}(\rho, \rho', \omega_k)$ is positive as long as $n(a_k) \leq \frac{N}{2}$.

6. $n(a_{k+t}) > n(a_k) = 0$; in this case the expression $\hat{\phi}(\rho, \rho', \omega_{k+t}) - \hat{\phi}(\rho, \rho', \omega_k)$ is positive as long as $n(a_{k+t}) \leq \frac{N}{2}$.

and thus $\hat{\phi}(\rho, \rho', \omega_k)$ is increasing in ω, and hence by the definition of first-order stochastic dominance, the entire expression in equation (33) is weakly positive and we are done.

$\square$

References

[1] Alonso, Ricardo and Odilon Camara. (2016a). "On the value of persuasion by experts". Working paper. Available at http://www-bcf.usc.edu/ ocamara/PersuasionExperts.pdf

[2] Alonso, Ricardo and Odilon Camara. (2016b) "Persuading Voters". Forthcoming, American Economic Review.

[3] Ricardo Alonso and Odilon Camara. (2016c). "Bayesian Persuasion with Heterogeneous Priors", Journal of Economic Theory, Volume 165, Pages 672-706.

[4] Arieli, Itai and Yakov Babichenko. (2016). "Private Bayesian Persuasion". Working paper.

[5] Aumann, Robert J, and Maschler, Michael B. (1995). Repeated Games with Incomplete Information. MIT Press, Cambridge, MA.

[6] Banks, Jeffrey S. and Sobel, Joel. (1987). "Equilibrium Selection in Signaling Games," Econometrica, Econometric Society, vol. 55(3), pages 647-61, May.

[7] Blackwell, David. (1951). "Comparison of Experiments." Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 93–102, University of California Press, Berkeley, Calif.

[8] Blackwell, David. (1953). "Equivalent Comparisons of Experiments." Ann. Math. Statist. 24, no. 2, 265–272.

[9] Cho, In-Koo and Joel Sobel. (1990). "Strategic Stability and Uniqueness in Signaling Games". Journal of Economic Theory, 50, 381-413.

[10] Crawford, Vincent P., and Joel Sobel. (1982). "Strategic Information Transmission." Econometrica, vol. 50, no. 6, pp. 1431-1451.

[11] Degan, Arianna and Li, Ming. (2015). "Persuasive Signalling" Working paper, available at SSRN: http://ssrn.com/abstract=1595511 or http://dx.doi.org/10.2139/ssrn.1595511

[12] Farrell, Joseph. (1993). "Meaning and Credibility in Cheap-Talk Games", Games and Economic Behavior, Volume 5, Issue 4, Pages 514-531

[13] Gentzkow, Matthew and Emir Kamenica. (2017a). "Competition in Persuasion." Review of Economic Studies, vol. 84, no. 4, pp. 300-322.

[14] Gentzkow, Matthew and Emir Kamenica. (2017b). "Bayesian persuasion with multiple senders and rich signal spaces." Games and Economic Behavior, Volume 104, Pages 411-429, https://doi.org/10.1016/j.geb.2017.05.004.

[15] Gill, David and Daniel Sgroi. (2012). "The optimal choice of pre-launch reviewer". Journal of Economic Theory, 147(3), 1247/1260

[16] Grossman, Sanford. (1981). "The Informational Role of Warranties and Private Disclosure about Product Quality", Journal of Law and Economics, 24, issue 3, p. 461-83.

[17] Grossman, Sanford J., Motty Perry. (1986). "Perfect Sequential Equilibrium", Journal of Economic Theory, Volume 39, Issue 1, Pages 97-119.

[18] de Groot Ruiz, Andrian, Theo Offerman and Sander Onderstal. (2013). "Equilibrium Selection in Cheap Talk Games: ACDC Rocks When Other Criteria Remain Silent." Working paper.

[19] Hedlund, Jonas. (2016). "Bayesian persuasion by a privately informed sender". Journal of Economic Theory, Volume 167, January 2017, Pages 229-268, ISSN 0022-0531, https://doi.org/10.1016/j.jet.2016.11.003.

[20] Kamenica, Emir, and Matthew Gentzkow. (2011). "Bayesian Persuasion." American Economic Review, 101(6): 2590-2615.

[21] Kohlberg, Elon, and Jean-Francois Mertens. (1986). "On the Strategic Stability of Equilibria." Econometrica, vol. 54, no. 5, pp. 1003-1037.

[22] Kreps, David M and Wilson, Robert. (1982). "Sequential Equilibria," Econometrica, Econometric Society, vol. 50(4), pages 863-94, July.

[23] Lehmann, E. L. (1988). "Comparing Location Experiments." The Annals of Statistics, vol. 16, no. 2, pp. 521-533.

[24] Mailath, George J., Masahiro Okuno-Fujiwara, Andrew Postlewaite. (1993) "Belief-Based Refinements in Signalling Games", Journal of Economic Theory, Volume 60, Issue 2, Pages 241-276.

[25] McKelvey, Richard D. and Thomas R. Palfrey. (1995). "Quantal Response Equilibria for Normal Form Games." Games and Economic Behavior, Volume 10, Issue 1, Pages 6-38, ISSN 0899-8256, http://dx.doi.org/10.1006/game.1995.1023.

[26] Milgrom, Paul, (1981). "Good News and Bad News: Representation Theorems and Applications", Bell Journal of Economics, 12, issue 2, p. 380-391.

[27] Milgrom, Paul and Shannon, Chris. (1994). "Monotone Comparative Statics," Econometrica, Econometric Society, vol. 62(1), pages 157-80, January.

[28] Myerson, R. (1978). "Refinements of the Nash Equilibrium Concept." International Journal of Game Theory, vol. 7, pp. 73-80.

[29] Osborne, Martin J. and Ariel Rubinstein. (1994). A Course in Game Theory. MIT Press, Cambridge, MA.

[30] Perez-Richet, Eduardo. (2014). "Interim Bayesian Persuasion: First Steps". American Economic Review: Papers & Proceedings, 104, 5.

[31] Persico, Nicola. (2000). "Information Acquisition in Auctions". Econometrica, vol. 68(1), pp. 135-148.

[32] Rayo, L., and Segal, I. (2010). "Optimal Information Disclosure". Journal of Political Economy, 118(5), 949-987.

[33] Selten, R. (1975) "Reexamination of the perfectness concept for equilibrium points in extensive games". International Journal of Game Theory 4: 25. doi:10.1007/BF01766400

[34] Taneva, Ina. (2016). "Information Design." Working paper, available at https://sites.google.com/site/itaneva13/research