Constrained Persuasion with Private Information

Andrew Kosenko*

September 12, 2021

Abstract

I study a model of strategic communication between a privately informed sender who can persuade a receiver using Blackwell experiments. Hedlund (2017) shows that private information in such a setting results in extremely informative equilibria. I make three points: first, the informativeness of equilibria relies crucially on two features - the mere availability of a fully revealing experiment, and a compact action space for the receiver. I show by examples that absent these features, equilibria may be uninformative. Secondly, I characterize equilibria in a simple model with constraints for the sender (only two experiments available, none are fully revealing) and the receiver (discrete action space). I argue that noisy experiments and discrete actions are the norm rather than the exception (and therefore, private information need not result in information revelation). Thirdly, I define a novel refinement that selects the most informative equilibria in most cases.

JEL Classification: D82, D83, C72.

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

*Assistant Professor of Economics, Department of Economics, Accounting, and Finance, School of Management, Marist College, kosenko.andrew@gmail.com. I would like to thank extremely helpful anonymous referees, Yeon-Koo Che, Ambuj Dewan, Vijay Krishna, Nate Neligh, Luca Rigotti, Burkhard C. Schipper, Jose Scheinkman, Marciano Siniscalchi, Joseph Stiglitz, Roee Teper, Richard van Weelden, the participants of the economic theory colloquium at Columbia for useful discussions and comments, and most of all, I am very grateful to my advisor, Navin Kartik, for his continued support, help and encouragement. Any errors are, of course, mine. Earlier versions of this paper were circulated under the titles "Bayesian Persuasion with Private Information" and "Noisy Bayesian Persuasion with Private Information"; the latter contains some extensions referred to in this paper.

1 Introduction

To what extent can one agent persuade another, less informed, agent to act by providing them with information? Is such persuasion ever credible, and if so, how much information can be conveyed in such a setting? How do the agents fare with regard to welfare? With mutual uncertainty about the payoff-relevant state of the world, the problem of information design with private information on one side has a number of intriguing features - not to mention the myriad applications.

I study these questions in a setting of a communication game with persuasive signaling. There is a single sender and a single receiver who share a commonly known prior belief about an unknown state of the world. The sender obtains a private, imperfectly informative signal about the state of the world, and armed with that knowledge[1] chooses an information structure - a Blackwell experiment (Blackwell (1951), (1953)) - that will generate a signal correlated with the state. All experiments have the same cost: zero. The receiver then has to take an action, based on the prior belief, the choice of experiment, as well as the realization of the signal from the experiment, that will affect the payoffs of both parties.

This paper is the first to study *constraints* on this problem. For the sender these constraints take the form of limiting the informativeness or number of experiments available, while for the receiver they take the form of a coarse action space. I show by examples that if the informativeness of experiments is bounded from above, or if the action space is coarse, the equilibria may be quite uninformative - a conclusion which stands in sharp contrast to earlier work. I then summarize the main features of equilibria of a model of constrained persuasion where there are only two experiments available. Concerned with two main issues - the informativeness of equilibria, and the welfare of the two sides, I apply a novel a refinement that typically selects the most informative equilibria.

The reason for introducing constraints is two-fold: first, they better reflect the actual actions available to the parties in applications. Secondly, the kinds of equilibria that are possible are very sensitive to the presence of such constraints - a problem with such constraints can have radically different, and uninformative, equilibria. The punchline is that with constraints, *private information matters*. In contradistinction to the unconstrained setting, the receiver may not find out the type of the sender or the state of the world in equilibrium in a model with constraints.

The setup is motivated by two important leading examples - a justice system setting where a district attorney is trying to persuade a grand jury to indict a defendant, and a drug approval setting where a pharmaceutical company is aiming to persuade a regulatory authority 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 district attorney, this may be something that the defendant had privately indicated to the counsel, or the attorney's past experience with similar cases; in the case of the pharmaceutical company this may be internal, preliminary, data or the views of scientists employed by the com-

[1] 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 (Alonso and Camara (2016)).

pany. In both cases the persuading party has to conduct a publicly visible experiment (a grand jury proceeding or a drug clinical trial) that may reveal something hitherto unknown to either party. A key feature of this setting is that the evidence, whether favorable to the attorney or the drug company, or not, from such an experiment cannot be concealed. In other words, once it is produced, the evidence cannot be hidden - but one may strategically choose not to produce it. Furthermore, I assume that evidence is produced stochastically - one can exercise only probabilistic control over the realizations of different pieces of evidence.

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 intends to capture the feature that there is nothing intrinsically different in the different types of senders and to isolate the effect of private information on outcomes. It will be useful to distinguish between three different receiver beliefs about the state of the world - ex-ante (before observing the choice of the experiment by the sender), ex-interim (after observing the choice of experiment, but before observing the experiment realization), and ex-post (after observing the experiment realization).

Hedlund (2017) studies a very closely related setup, finding that in a model where all experiments, including a fully revealing experiment (FRX, for brevity), are available, and a compact, convex, and connected action space for the receiver, *private information forces equilibria to be extremely informative*. There, equilibria are of two kinds - "either separating (i.e., the sender's choice of signal reveals his private information to the receiver) or fully disclosing (i.e., the outcome of the sender's chosen signal fully reveals the payoff-relevant state)" (Hedlund (2017) p. 1). In other words, the receiver either directly finds out the state, or at least he finds out the sender's type. In the first case the sender's private information is irrelevant. In the second it is always fully revealed. Thus, private information does not help the sender (in particular, the sender should never pay for private information), and in that sense, private information in that setting does not matter.[2] Among other results is the fact that "the sender prefers the symmetric information benchmark over the equilibrium," while "the receiver prefers the equilibrium (i.e., the receiver benefits from her ignorance)" (Hedlund (2017) p. 4).

This paper makes three points. The first is that the mere presence - regardless of whether it is chosen in a particular equilibrium or not - of an FRX is important. Equilibria in a model where an FRX is not available are different. If an FRX in unavailable (and an arbitrarily small amount of noise in only a single state is enough), private information need not result in very informative outcomes. Likewise, a model with a binary action space (even if all experiments are available) has relatively uninformative equilibria.

The results, absent an FRX, or a compact action space, are intriguing: for example, the sender is not indifferent across all equilibria as in Hedlund (2017), and may or may not benefit from

[2] Alonso and Camara (2018) show that if a fully revealing information structure is available, then an uninformed sender 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.

being privately informed, relative to the symmetric information benchmark. The receiver can still benefit from persuasion, even in this setting of private information and known "ulterior" motives on the part of the sender.

Aside from the contrast with the work of Hedlund (2017) vis-à-vis the *informativeness* of equilibria, there is also an interesting parallel with Alonso and Camara (2018) with regard to *welfare*. In a somewhat different context they show that in the presence on an FRX, the sender does not benefit from private information, while in its absence the welfare of the sender is ambiguous. I reach the same conclusion, studying a more parsimonious and explicit model, with full equilibrium characterization. Thus, the presence of an FRX is important not only for the kinds of equilibria, and their informativeness, but also for their welfare properties. The present paper is not, however, a special case of Alonso and Camara (2018)[3], which further underscores the significance of the FRX for interpretation of the results of models in this literature.

The reason that a compact, connected, and convex action space for the receiver is important in the present model is that without it - with any "coarseness" - the action of the receiver will be locally constant for some ranges of the posterior belief. This results in a lack of strict monotonicity of the payoffs, which in turn, makes possible the uninformative equilibria. Of course, in most applications, actions are, indeed, coarse: approve/deny, convict/acquit, and so on.

The reason that the FRX is important is because it provides a possible deviation for the sender with the (very special) property that upon choosing it, there is no flexibility about receiver interim beliefs. Since the experiment is fully revealing about the state, interim beliefs are irrelevant. The fact that this action and its consequences are always available means that in any equilibrium, no type of sender can prefer to deviate to it. This, however, eliminates equilibria where types pool on less informative experiments; these equilibria are ruled out in a model with an FRX and a compact action space because if they were to exist, the "good" type of sender would always (under reasonable assumptions on preferences) wish to distinguish herself, and deviate to the fully revealing experiment.[4] The sender is, in a sense, guaranteed a "safe" (though fully revealing) action that does not depend on interim beliefs. If an FRX is absent, whether the good type wants to deviate (to a "most informative" experiment) is less clear, because there is now the question of what receiver interim beliefs are upon observing such a deviation, and for some beliefs, the receiver may not choose the action that is preferred by the sender, and therefore, the sender might not wish to deviate to a most informative experiment in the first place. This makes the existence of pooling equilibria (where the pooling is on a less informative experiment) possible.

[3]Using the language of Alonso and Camara (2018), the experiments studied here are not "redundant" - observing the outcome of an additional experiment will generate more information, and change the beliefs of the receiver.

[4]For intuition for this result note that in a pooling equilibrium the receiver's interim belief must be the prior. The sender prefers high actions; in equilibria where the receiver follows the signal realization this means that on path, the probability of the good signal realization (and thus the good action) is proportional to the interim belief (in this case, the prior) and the probability of the good signal realization. But the beliefs of the good type of sender are different from the prior - from her point of view, the good state is more likely, and since the fully revealing experiment always reveals the state, the probability of a good signal realization - from the point of view of the good type of sender - is always higher than that in a pooling equilibrium. Note that this argument depends on the interim beliefs being irrelevant when the sender chooses the FRX.

Bounding the informativeness of experiments from above, thus constraining the action space for the sender (example 1), or introducing coarseness in receiver actions (example 2), or doing both (example 3) changes the outcome dramatically. Hence an economist should (perhaps) care about what may seem to be a technicality - the question of whether extremely informative experiments are available is important because its mere presence, regardless of whether it is used or not, will determine the informational and welfare properties of the outcome. In any reasonable application an FRX is typically unavailable, which of course implies, that the relevant kinds of equilibria are of the kind studied here.[5] The very same is true of a compact action space.

This discussion is graphically summarized in Table 1; in gray are highlighted the combinations of assumptions studied in the paper, with corresponding numbered examples.

	Compact action space for R	Discrete action space for R
FRX available	Either separating or fully revealing equilibria (Hedlund (2017))	Uninformative outcomes possible Example 2
FRX unavailable	Uninformative outcomes possible Example 1	Uninformative outcomes possible Example 3

Table 1: Summary of Assumptions and Possible Equilibria

Given these disquieting conclusions - that the previous results on private information leading to very informative equilibria - relied on assumptions that are unlikely to be satisfied, and that without these assumptions, equilibria may be uninformative, is there any way of reobtaining the (very desirable) informative equilibria? What is a natural equilibrium refinement that has economic content, and makes a meaningful and attractive selection when other refinements are silent? That is the third and last contribution of this paper. The refinement (operating by restricting off-path beliefs), which I term "belief-payoff monotonicity", or BPM, relies on a particular illustrative kind of reasoning: it requires that upon observing a deviation, the receiver must believe that it is coming from the type that benefits *relatively* more from that particular deviation; this belief consistency requirement rules out some of the less informative equilibria.

The rest of the paper is organized as follows: in the next section, I discuss the literature and place the model in context. Section 3 describes in detail the setting, while sections 4, 5, and 6 contain examples 1, 2, and 3; section 6 also characterizes equilibria in a model of constrained persuasion with only two experiments for the sender and two actions for the receiver. Section 7 introduces and applies the BPM refinement, discusses welfare, and briefly concludes. Generalizations that show that the results are robust to expanding the set of information structures and non-dichotomous states of nature can be found in Kosenko (2021). The proof of proposition 7 is in the appendix.

[5]Interestingly, in work on multi-sender persuasion (Gentzkow and Kamenica (2017a), Gentzkow and Kamenica (2017b)) a similar insight has emerged - the capability of one player to unilaterally mimic a particular distribution of signals ("Blackwell-connectedness", which can be thought of as an analogue to a fully revealing experiment in a single-sender framework) has become a key condition.

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 "Bayesian persuasion"; they also identify conditions under which the sender "benefits from persuasion", utilizing a "concavification" technique introduced in Aumann and Maschler (1995).

Hedlund (2017) is very closely related work; he works with a similar model (and also uses a refinement concept - D1) 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. I 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 I consider a simpler model where one can explore the role of particular assumptions and show the importance of these features for equilibrium welfare. In particular, I consider experiments where a fully revealing experiment is not available, and the action space for the receiver is not a compact interval; these assumption appears to be more realistic in applications and create an additional level of difficulty not present in Hedlund (2017).

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 the (many) equilibria 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 (2018) show that with an FRX, the sender can not benefit from becoming an expert (i.e. from learning some private information about the state). Among other results, they provide a condition ("redundancy" - the ability to duplicate a distribution of signal realizations) under which the sender can never benefit from becoming informed. In the present setting this condition will not be satisfied, and indeed, I find that the sender can sometimes benefit. They work with a more general model but make several assumptions to obtain sharp characterizations. In contrast, I work with a very simple but explicit model that is perhaps more illuminating.

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 (2016) present a similar model 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 (2016) endow their senders with state-dependent utility functions.

3 Model

Setup

Consider a strategic communication game between a sender (she) and receiver (he), where the sender (S) is privately, though imperfectly, informed about the state of the world. Consequently, the receiver (R) will form beliefs about both the type of the sender and the state of the world.

There is an unknown state of the world, $\omega \in \Omega = \{\omega_L, \omega_H\} = \{0, 1\}$, unknown to both parties with a commonly known prior probability of $\omega = \omega_H$ equal to $\pi \in (0, 1)$. The assumption of a binary state is maintained throughout the paper (see Kosenko (2021) for extensions and additional discussion). The sender can be one of two types: $\theta \in \Theta = \{\theta_L, \theta_H\}$. The sender's type is private information to her. The type structure is generated as follows:

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

for $\xi \geq \frac{1}{2}$, with the other signals occurring with the complementary probability.

A key feature distinguishing this model from others is that the private information of the sender is not about her preferences, but about the state of nature. In this sense the sender is more informed than the receiver. By contrast, in, say, Perez-Richet (2014), the state is identified with the preference parameter.

The sender chooses an *experiment* - a complete conditional distribution of *signals* given states; all experiments have the same cost, which I set to zero. The choice of the experiment and the realization of the signal are observed by both the sender and the receiver. The set of available experiments is $\mathbf{\Pi}$ with typical element Π. The sets $\mathbf{\Pi}$ will be different in examples 1, 2, and 3.

The receiver takes an action $a \in A$; the action sets will also vary in the examples. Let $\mu(\Pi) = \mathbb{P}(\omega = \omega_H | \Pi)$ 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. Let $\beta(\omega_H | \Pi, \sigma, \mu)$ be the *posterior* belief of the receiver that the state is high conditional on observing Π and σ, given interim beliefs μ. 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.

Denote by $p(\theta) \in \Delta(\Pi)$ the strategy of the sender, let $v(\Pi, \theta, q) \triangleq \mathbb{E}\left(u^S(a) | \Pi, \theta, q\right)$ be the expected value of choosing experiment Π for a sender of type θ when the receiver uses a strategy $q(\Pi, \mu) \in \Delta(A)$, and $\hat{v}(\Pi_i, \mu, \theta_j) \triangleq \mathbb{E}_{\sigma,a}(u^S(a) | \Pi_i, \mu)$ denote the expected value of choosing an experiment Π_i for type θ_j when the receiver's interim beliefs are exactly μ. The function $\hat{v}$ is piecewise linear in μ and continuous in the choice of the experiment.

Equilibrium Concept

The basic equilibrium concept is PBE:

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

1. *Sequential Rationality:*

$$\forall \theta, p(\theta) \in arg\ max\ v(\Pi, \theta, q)\ and\ a^*(\Pi, \sigma) \in \arg\max \sum_{\omega} u(a, \omega)\beta(\omega|\Pi, \sigma) \tag{2}$$

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 first two parts of the definition are standard. I augment the definition with a tie-breaking rule (the third requirement) to facilitate the exposition. The rule requires that whenever the receiver is indifferent between two actions, he always chooses the one preferred by the sender.[6] 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.

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.
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.

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

4 Example 1: A Compact Action Space and An Upper Bound on Informativeness

This section exhibits an example showing that bounding the informativeness of experiments from above changes the kind of equilibria that can arise. The example shares some features with the canonical example of KG, and some features with Hedlund (2017); essentially I am interested in the kinds of equilibria that can arise in the environment studied by Hedlund (2017), but with a constraint on the informativeness of experiments.

Let the common prior be $\pi = \frac{3}{10}$, let $\theta_L = \frac{1}{1000}, \theta_H = \frac{99}{100}$, the action space for the receiver be $A = [0,1]$, and utilities given by $u^S(\omega, a) = a, u^R(\omega, a) = -(\omega - a)^2$.[7]

Experiments are represented as matrices, with the (i,j)'th entry reflecting the probability of signal i in state j.

$$\Pi = \begin{matrix} \\ \sigma_L \\ \sigma_H \end{matrix} \begin{matrix} \begin{matrix} \omega_L & \omega_H \end{matrix} \\ \begin{pmatrix} \rho_0 & 1-\rho_1 \\ 1-\rho_0 & \rho_1 \end{pmatrix} \end{matrix}$$

with $\rho_0, \rho_1 \in \left[\frac{1}{2}, \overline{\rho}\right]$. The sender is free to choose any ρ_0 and ρ_1 in this set. Of course, if $\overline{\rho} = 1$, a fully revealing experiment is available, and is simply the 2×2 identify matrix denoted by Π^{FI}. To capture the fact that a fully revealing experiment is not available, put an upper bound on the informativeness of the signals, say $\overline{\rho} = 0.9$, and refer to the maximally (as opposed to fully) informative experiment as $\Pi^{MI} = \begin{pmatrix} \frac{9}{10} & \frac{1}{10} \\ \frac{1}{10} & \frac{9}{10} \end{pmatrix}$.

Say that Π' is *more informative* than Π if $\rho_0' \geq \rho_0$ and $\rho_1' \geq \rho_1$. To make the contrast with Hedlund (2017) as stark as possible, let us also use the same refinement - D1 (Kohlberg and Mertens (1986), Cho and Sobel (1990)). Given a sender strategy $p(\theta)$ and receiver interim beliefs $\tilde{\mu}$, define for $\Pi \in \mathbf{\Pi}$ and $\theta_b \in \{\theta_L, \theta_H\}$ the sets $D^0(\Pi, \theta_b) \triangleq \{\mu \in [\theta_L, \theta_H] \,|\hat{v}(\Pi, \mu, \theta_b) \geq \hat{v}(p(\theta_b), \mu(\tilde{p}(\theta), \theta_b)\}$ and $D(\Pi, \theta_b) \triangleq \{\mu \in [\theta_L, \theta_H] \,|\hat{v}(\pi, \mu, \theta_b) > \hat{v}(p(\theta_b), \mu(\tilde{p}(\theta), \theta_b)\}$. That is, fixing an equilibrium and associated utility levels, D^0 and D are the sets of receiver interim beliefs such that type θ_b of sender weakly ($D^0(\Pi, \theta_b)$) and strictly ($D(\Pi, \theta_b)$) benefits from deviating to an experiment π, provided that the receiver best-responds using beliefs μ defined by $D^0(\Pi, \theta_b)$ and $D(\Pi, \theta_b)$. An equilibrium survives D1 if $D^0(\Pi, \theta_b) \supsetneq D(\Pi, \theta_i)$ implies that $\mu(\Pi) = \theta_i$.

Following Hedlund (2017), call a PBE that survives D1 criterion is a *D1 equilibrium*. The following proposition is the first contribution of the paper, and shows that with an upper bound on the informativeness of experiments, equilibria may be relatively uninformative - the state of the world or the type of the sender are not revealed in this equilibrium.

Proposition 1. *Let* $\mathbf{\Pi}_{\frac{9}{10}} \triangleq \{\Pi | \frac{1}{2} \leq \rho_0, \rho_1 \leq \overline{\rho} = \frac{9}{10}\}$ *and* $\Pi^p = \begin{pmatrix} \frac{4}{7} & \frac{1}{10} \\ \frac{3}{7} & \frac{9}{10} \end{pmatrix}$. *The pair of strategies*

[7] This specifications satisfies all the assumptions made by Hedlund (2017).

$p(\theta_H) = p(\theta_L) = \delta(\Pi^P)$, *where* $\delta(x)$ *is the Dirac delta distribution, with* $\mu(\Pi^P) = 0.3, \mu(\Pi) = \pi$ *for any* $\Pi \neq \Pi^P, \Pi \in \mathbf{\Pi}_{\frac{9}{10}}$, *is a D1 equilibrium.*

Π^p is the optimal experiment for the sender of type $\frac{3}{10}$, *à-la* KG.

Proof. To verify this compute directly: $\hat{v}(\Pi^p, \mu = \pi, \theta_H) \approx 0.4143$, and $\hat{v}(\Pi^p, \mu = \pi, \theta_L) \approx 0.0029$. If either type of the sender deviates to Π^{MI} (which is also the experiment maximizing the expected payoff of the high type, provided that the receiver attributes it to the low type), they obtain $\hat{v}(\Pi^{MI}, \mu = \theta_L, \theta_H) \approx 0.0080$ and $\hat{v}(\Pi^p, \mu = \theta_L, \theta_L) \approx 0$; neither deviation is profitable, provided the receiver interprets the deviation as coming from the low type ($\mu = \theta_L = 0.001$).

To verify that this equilibrium survives D1, compute the relevant sets: $D(\Pi^{MI}, \theta_H) = (\approx 0.1087, 1]$ and $D(\Pi^{MI}, \theta_L) = (\approx 0.0027, 1]$. As such $D^0(\Pi^{MI}, \theta_L) \supsetneq D(\Pi^{MI}, \theta_H)$, and the D1 criterion stipulates that the receiver should, in fact, believe that the deviation is coming from the low type. In other words, this equilibrium survives D1. □

Notably, if Π^{FI} were available, the high type of sender would be able to secure a payoff of $\frac{99}{100} \times 1 = 0.99$, thus yielding a profitable deviation for the high type. In other words, the equilibrium when an FRX is unavailable may be only imperfectly informative, and private information matters.

5 Example 2: A Coarse Action Space and All Experiments Available

The same conclusion obtains if all experiments were available, but the action space is binary.[8] Say $A = \{0, 1\}$, the prior is $\pi = \frac{3}{10}$, and $\theta_L = \frac{2}{10}, \theta_H = \frac{4}{10}$ with $u^S(\omega, a) = a$ and $u^R(\omega, a) = -(\omega - a)^2$. As before, let Π^{FI} be the 2×2 identity matrix.

Proposition 2. *Let* $\mathbf{\Pi}^* \triangleq \{\Pi | \frac{1}{2} \leq \rho_0, \rho_1 \leq \overline{\rho} = 1\}$ *and* $\Pi^{KG} = \begin{pmatrix} \frac{4}{7} & 0 \\ \frac{3}{7} & 1 \end{pmatrix}$. *The pair of strategies* $p(\theta_L) = p(\theta_H) = \delta(\Pi^{KG})$, *where as before,* $\delta(x)$ *is the Dirac delta distribution, with* $\mu(\Pi^{KG}) = \pi, \mu(\Pi) = 0.2$ *for any* $\Pi \in \mathbf{\Pi}^*, \Pi \neq \Pi^{KG}, \Pi^{FI}$, *is a D1 equilibrium.*

Proof. The best deviation is $\Pi^{Dev} = \begin{pmatrix} \frac{6}{7} & 0 \\ \frac{1}{7} & 1 \end{pmatrix}$, provided that the receiver assigns $\mu = 0.2$ for any experiment off the equilibrium path. I once again compute directly: $\hat{v}(\Pi^{KG}, \mu = \pi, \theta_H) \approx 0.6571$, which is greater than $\hat{v}(\Pi^{Dev}, \mu = \theta_L, \theta = \frac{4}{10}) \approx 0.4857$. Similarly, $\hat{v}(\Pi^{KG}, \mu = \pi, \theta_L) \approx 0.5428$, which is greater than $\hat{v}(\Pi^{Dev}, \mu = \theta_L, \theta_L) \approx 0.3143$. This equilibrium also survives D1: $D^0(\Pi^{Dev}, \theta_H) = (\approx 0.6171, 1] \subsetneq D(\Pi^{Dev}, \theta_L) = (\approx 0.5229, 1]$. As before, private information matters - the receiver may not find out the type of sender she is dealing with, or the state of the world, and the equilibrium may be uninformative relative to the Hedlund (2017) benchmark. □

[8] This example was suggested by an anonymous referee, whose contribution I gratefully acknowledge and highlight.

6 Example 3: A Coarse Action Space with Two Experiments

Turning now to the conjunction of the two kinds of constrains, I show that the same kinds of relatively uninformative equilibria persist in a setting with finitely many actions for the receiver and an upper bound on informativeness. In this section I study a parsimonious model of such a setting, characterizing all equilibria, and summarizing their main features. To illustrate the main insights as sharply as possible, I constrain the sender to choose among only two experiments: $\Pi_H = \begin{pmatrix} \rho_H & 1-\rho_H \\ 1-\rho_H & \rho_H \end{pmatrix}$ and $\Pi_L = \begin{pmatrix} \rho_L & 1-\rho_L \\ 1-\rho_L & \rho_L \end{pmatrix}$ with $\rho_H > \rho_L$ so that Π_H is more informative than Π_L. I assume (naturally) that the experiment realizations are independent of the realization of the sender's type. The available actions for the receiver are $a \in A = \{a_H, a_L\}$.

In this model, there are several classes of equilibria: a unique separating equilibrium where the high type chooses the more informative experiment, and several continua of pooling equilibria where the pooling can be on either experiment. It is the possibility of pooling on the less informative (arbitrarily uninformative, in fact) experiment that is among the surprising features: this outcome is not possible in a model with all experiments available and a compact action space.

6.1 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. Conceretly, suppose that $u^S(a_H) = 1$, $u^S(a_L) = 0$, and the receiver has preferences given by $u^R(\omega_H, a_H) = 1, u^R(\omega_H, a_L) = -1, u^R(\omega_L, a_L) = 1, u^R(\omega_L, a_H) = -1$. The symmetry in the payoffs is special, but doesn't affect the qualitative properties of equilibria. Importantly, there is no single-crossing assumption on the primitives in this model. Rather, a similar kind of feature is derived endogenously.

6.2 Perfect Bayesian equilibria

It will be convenient in this section 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.

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

$$a^*(\Pi, \sigma) \in \arg\max_{\Delta\{a_H, a_L\}} u^R(a, \omega_H)\beta(\Pi, \sigma) + u^R(a, \omega_L)(1 - \beta(\Pi, \sigma)) \tag{3}$$

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

As a first step let us 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,

$$\forall \theta, p(\theta) = 1 \iff v(\Pi_H, \theta, q) > v(\Pi_L, \theta, q) \tag{4}$$

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 usual, in evaluating the observed signal the receiver uses a conjecture of the sender's strategy, correct in equilibrium. If an FRX 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.

In what follows I focus on the 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.[9] To that end, consider

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.*

It is immediate that if an equilibrium is nontrivial, it is also p-nontrivial, but not vice versa. From now on I will focus only on (p-)nontrivial equilibria;[10] this amounts to placing restrictions on the four parameters that I will be explicit about when convenient. This does not cover all possible equilibria for all possible parameters, but it does focus on the "interesting" equilibria, where the action of the receiver depends on the realized signals.

There are four key classes of equilibria, key in the sense that are important for interpreting the qualitative conclusions of the model, and are therefore, economically significant.[11] See Kosenko (2021) for a complete list of possible classes of equilibria, and conditions for their existence.

[9] 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. I 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.

[10] 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 I do not consider such equilibria.

[11] These equilibria 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 equi-

First, there is a unique separating equilibrium, in which the high type of sender chooses the more informative experiment, and the low type chooses the less informative one:[12]

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 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.

There are two kinds of fully nontrivial equilibria - one where both types pool on Π_H, the more informative experiment (prop. 4), and one where they pool on Π_L (prop. 5):

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" for "fully nontrivial with pooling on the highly informative experiment".*

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" for "fully nontrivial with pooling on the less informative experiment".*

There exists yet another kind of equilibria with pooling on the Π_L - a p-nontrivial one. The nomenclature for the p-nontrivial equilibria is as follows:

$$\underbrace{PNT}_{\text{Equilibrium type}} - \underbrace{L}_{\text{On-path action}} \underbrace{H}_{\text{If R sees this}} (\underbrace{a_L}_{\text{R takes this action}}) \tag{5}$$

Proposition 6. *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)".*

It is the fact that equilibria of the kind described in propositions 5 and 6 can exist that is among the key lessons of studying a model where fully revealing experiments are unavailable. It turns out that there are seven classes kinds of equilibria in total: $SEP, FNT-H, FNT-L, PNT-LH(a_L), PNT-HL(a_L), PNT-HH(a_H), PNT-LL(a_H)$.[13] To summarize, there is a unique sep-

librium path, and verifying best responses, using Bayes rule whenever possible. I omit the tedious but straightforward computations. For convenience, for any variable $x \in (0,1)$ denote by $\tilde{x}$ the ratio $\frac{x}{1-x}$.

[12] The "information validates the prior", or IVP theorem of Kartik, Lee, and Suen (2020) also immediately reveals this equilibrium outcome.

[13] There is yet another subtlety that arises - because not all of these kinds of equilibria exist for all parameters, one might ask whether they co-exist. If they did not, the question of multiple equilibria would, perhaps, be moot. The equilibria do co-exist in the relevant cases (Kosenko (2021)).

arating equilibrium, and fully pooling and p-nontrivial pooling equilibria where the pooling can be on either experiment.

7 Refinement: BPM Criterion

Often, when there is a problem of multiplicity of equilibria, equilibrium refinements are used to select among them. Hedlund (2017) uses PBE augmented with the D1 criterion (Cho and Sobel (1990)); here this particular refinement does not help.[14] Other standard refinements for signaling games such as perfect sequential equilibria (Grossman and Perry (1986)), neologism-proof equilibria (Farrell (1993)),[15] perfect (Selten (1975)), or proper (Myerson (1978)) equilibria, also do not narrow down predictions, for similar reasons.

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 PBE in this game can be sequential equilibria. Undefeated equilibrium still does not go far enough, as I discuss in Kosenko (2021).

Yet, not all is lost. 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: while neither type benefits 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.[16] 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

[14] It can be checked by direct computation that all of the equilibria described above survive criterion D1. 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. This is due to the fact that for all equilibria and deviations, criterion D1 requires a *strict* inclusion of the D sets, while in this game the relevant D sets are, in fact, identical for both types. Similarly, other related refinements such as the intuitive criterion (Cho and Kreps (1987)) do not help (IC does not work because for the right range of beliefs both types benefit. Note also that were this not true, I would be in the range of parameters where the separating equilibrium occurs - c.f. SEP.) and other refinements based on strategic stability (Kohlberg and Mertens (1986)). 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 I 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 I cannot support all equilibria by mixing.

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

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

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 *b*elief-*p*ayoff *m*onotonicity. I now turn to this refinement, and show that it does help narrow down the predictions to some degree. I give an ordinal definition that is suitable to the present environment, but it can be generalized in a straightforward way.

Definition 4 (BPM Criterion)**.** *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. I scale the differences in a way that makes the definition invariant to strictly increasing transformations of the sender's payoffs (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. Finally, if an equilibrium does not fail the BPM criterion, I say that it survives it. From now on I will refer to a PBE with tie-breaking that also survives criterion BPM as a *BPM equilibrium*. The last result of the paper is:

Proposition 7. *The following classes of equilibria are BPM equilibria: SEP, FNT-H, FNT-L, PNT-HL(a_L), PNT-HH(a_H) and PNT-LL(a_H).*

It should be noted that these equilibria are also ϵ-BPM equilibria, for all admissible ϵ, but I suppress this fact in the exposition that follows. Interestingly, BPM does not help eliminate the FNT-L equilibrium, but that is because the only case in which it coexists with FNT-H is the knife-edge case where $\pi = \xi = \frac{1}{2}$, so that the private signal is uninformative, the utilities of the high and low type are identical in both equilibria, and both types are exactly indifferent in between following their equilibrium strategy or deviating to a more informative experiment. The intuition for why PNT-LH(a_L) is ruled out is illustrated in figure 1:

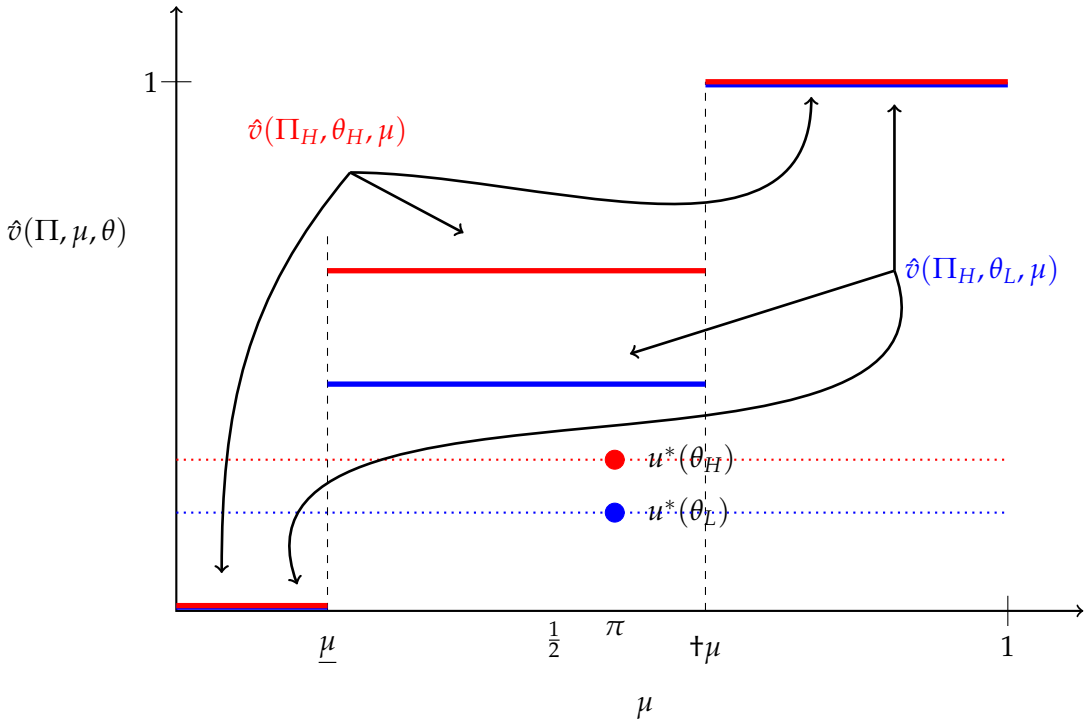

Figure 1: Illustration with pooling on Π_L, and the deviation to Π_H.

In figure 1 the horizontal doted lines represent the on-path,[17] equilibrium utility levels in the PNT-LH(a_L) 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 utility of deviating from the equilibrium path in negative in this equilibrium, since upon observing an off-path deviation the receiver's belief is $\mu < \underline{\mu}$. 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 $\mu \in [0, \underline{\mu})$[18] 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

[17] Here an throughout I 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. I do not use terms like "out of equilibrium" since that could create confusion.

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

such a deviation for those beliefs. The crucial region is $\mu \in [\underline{\mu}, \dagger\mu)$. It is here that criterion BPM operates efficiently - both types get positive payoff from the equilibrium and the deviation, but I have shown above that the high type benefits relatively more. Beliefs above $\mu\dagger$, again, cannot sustain a nontrivial equilibrium, and hence there is no need to consider them as 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 $\mathbb{P}(\sigma_H|\theta_H) > \mathbb{P}(\sigma_H|\theta_L)$. Nevertheless, given the restrictions on parameter discussed above, BPM does, in fact eliminate the p-nontrivial 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. No reasonable refinement could ever refine that outcome away, since the sender would never profitably deviate from that equilibrium. As mentioned above, undefeated equilibrium does help to refine predictions, however, and in fact, makes a very similar selection.

7.1 Welfare and Comparative Statics: Do The Players Benefit from Persuasion with Private Information?

We now turn to the question of welfare. For the receiver[19], 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. 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, it is clear that the receiver strictly prefers the outcome under persuasion to 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, 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,

[19]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.

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:

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

Figure 2: Sender Preferences Over Equillibria

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.[20] Notably, this is quite starkly different to the results of Hedlund (2017), 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.

A natural question 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 $\pi \geq \frac{1}{2}$, and the 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 the sender sometimes benefits from private information. This is in in line with Alonso and Camara (2018) 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 (2018) make precise; this feature also allows an informed sender to be better or worse off. Furthermore, here the sender does not benefit from persuasion[21] (and in fact does strictly worse), if the receiver is ex-ante willing to take the high action (if $\pi \geq \frac{1}{2}$), and does strictly better otherwise. This observation has an analogue in KG: there, the sender benefits if the receiver is willing ex-ante to take an action that is inferior from the point of view of the sender.

[20] Again, with the caveat that FNT-L and FNT-H coexist in a knife-edge case.

[21] 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.

References

[1] Alonso, Ricardo and Odilon Camara. (2018). "On the value of persuasion by experts". Journal of Economic Theory, vol. 174, pp. 103-123.

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

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

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

[5] 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.

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

[7] Cho, In-Koo and David M. Kreps. (1987). "Signaling games and stable equilibria". Quarterly Journal of Economics 102, pp. 179-221.

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

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

[10] 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

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

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

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

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

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

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

[17] 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.

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

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

[20] Kartik, Navin, Frances Xu Lee, and Wing Suen. (2020). "Information Validates the Prior: A Theorem on Bayesian Updating and Applications". July, forthcoming in American Economic Review: Insights.

[21] Kohlberg, Elon, and Jean-Francois Mertens. (1986). "On the Strategic Stability of Equilibria." Econometrica, vol. 54(5), pp. 1003-1037.

[22] Kosenko, A. (2021). "Noisy Bayesian Persuasion with Private Information". mimeo.

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

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

[25] 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.

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

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

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

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

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

[31] Perez-Richet, Eduardo. (2014). "Interim Bayesian Persuasion: First Steps". American Economic Review: Papers & Proceedings, vol. 104(5), pp. 469-474.

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

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

[34] 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

[35] Spence, Michael. (1973). "Job Market Signaling". Quarterly Journal of Economics, vol 87, issue 3, pp. 355-374.

Appendix

Proof of Proposition 7. 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 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, $u^*(\theta_H) =$

$$\begin{aligned}\hat{v}(\Pi_L, \pi, \theta_H) = \rho_L \left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_L)\left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right]\end{aligned} \tag{6}$$

and $u^*(\theta_L) =$

$$\begin{aligned}\hat{v}(\Pi_L, \pi, \theta_L) = \rho_L \left[\mathbb{P}(\omega_H|\theta_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_L)\left[\mathbb{P}(\omega_H|\theta_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right]\end{aligned} \tag{7}$$

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

$$\begin{aligned}\hat{v}(\Pi_H, \mu, \theta_H) - u^*(\theta_H) = \rho_H \left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{|\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_H)\left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}}\right] - \\ -\rho_L\left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}}\right] + \\ +(1-\rho_L)\left[\mathbb{P}(\omega_H|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}} + \mathbb{P}(\omega_L|\theta_H)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right] = \\ = (\mathbb{P}(\omega_H|\theta_H))\left[\rho_H\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} - \rho_L\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}} + (1-\rho_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}} - \right. \\ \left. -(1-\rho_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}}\right] + (\mathbb{P}(\omega_L|\theta_H))\left[\rho_H\mathbb{1}_{|\{\mu|\beta(\Pi_i,\sigma_L,\mu)\geq\frac{1}{2}\}} - \rho_L\mathbb{1}_{\{\beta(\Pi_L,\sigma_L,\pi)\geq\frac{1}{2}\}} + \right. \\ \left. +(1-\rho_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} - (1-\rho_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right]\end{aligned} \tag{8}$$

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. $\overline{\mu} = 1-\rho_L$) 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, I focus on nontrivial equilibria (so that I disregard the terms that involve observing the low signal/action).

Now compute directly:

$$
\begin{aligned}
&\hat{v}(\Pi_H, \theta_H, \mu) - u^*(\theta_H) - (\hat{v}(\Pi_H, \theta_L, \mu) - u^*(\theta_L)) = \\
&= [\mathbb{P}(\omega_H|\theta_H) - \mathbb{P}(\omega_H|\theta_L)] \left[\rho_H \mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} - \rho_L \mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right] + \\
&+[\mathbb{P}(\omega_L|\theta_H) - \mathbb{P}(\omega_L|\theta_L)] \left[(1-\rho_H)\mathbb{1}_{\{\mu|\beta(\Pi_i,\sigma_H,\mu)\geq\frac{1}{2}\}} - (1-\rho_L)\mathbb{1}_{\{\beta(\Pi_L,\sigma_H,\pi)\geq\frac{1}{2}\}}\right] = \\
&= \begin{cases} u^*(\theta_L) - u^*(\theta_H) < 0, & \text{for } \mu \in [0, \underline{\mu}) \\ 2(\rho_H - \rho_L)(\mathbb{P}(\omega_H\theta_H) - \mathbb{P}(\omega_H|\theta_L))) > 0 & \text{for } \mu \in [\underline{\mu}, \dagger\mu) \\ 2\rho_L[\mathbb{P}(\omega_H|\theta_L) - \mathbb{P}(\omega_H|\theta_H)] + \mathbb{P}(\omega_H|\theta_H) - \mathbb{P}(\omega_H|\theta_L) < 0 & \text{for } \mu \in [\dagger\mu, 1] \end{cases}
\end{aligned}
\tag{9}
$$

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 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 6 and note that criterion BPM restricts beliefs off the equilibrium path) and we are done. □