Working Paper Series 708
(ISSN 1211-3298)

Form of Preference Misalignment Linked to State-Pooling Structure in Bayesian Persuasion

Rastislav Rehák
Maxim Senkov

CERGE-EI
Prague, October 2021

ISBN 978-80-7343-515-8 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-610-9 (Národohospodářský ústav AV ČR, v. v. i.)

Form of Preference Misalignment Linked to State-Pooling Structure in Bayesian Persuasion*

Rastislav Rehák† Maxim Senkov‡

CERGE-EI§

October 22, 2021

Abstract

We study a Bayesian persuasion model in which the state space is finite, the sender and the receiver have state-dependent quadratic loss functions, and their disagreement regarding the preferred action is of arbitrary form. This framework enables us to focus on the understudied sender's trade-off between the informativeness of the signal and the concealment of the state-dependent disagreement about the preferred action. In particular, we study which states are pooled together in the supports of posteriors of the optimal signal. We provide an illustrative graph procedure that takes the form of preference misalignment and outputs potential representations of the state-pooling structure. Our model provides insights into situations in which the sender and the receiver care about two different but connected issues, for example, the interaction of a political advisor who cares about the state of the economy with a politician who cares about the political situation.

Keywords: Bayesian persuasion, strategic state pooling, preference misalignment, graph procedure

---

*Thanks to Jan Zápal, Ole Jann, Inés Moreno de Barreda, Fedor Sandomirskiy, Ludmila Matysková, Filip Matějka, and the conference audience at GAMES 2020/1 for useful comments.

†Email: Rastislav.Rehak@cerge-ei.cz. Rastislav Rehák was supported by Charles University GAUK project No. 666420 and by the H2020-MSCA-RISE project GEMCLIME-2020 GA No. 681228. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreements No. 101002898 and No. 770652).

‡Email: msenkov@cerge-ei.cz. Maxim Senkov was supported by the Lumina Quaeruntur fellowship (LQ300852101, Challenges to Democracy) of the Czech Academy of Sciences.

§CERGE-EI, a joint workplace of Center for Economic Research and Graduate Education, Charles University and the Economics Institute of the Czech Academy of Sciences, Politických vězňů 7, P.O. Box 882, 111 21 Prague 1, Czech Republic.

1 Introduction

Bayesian persuasion, pioneered by Kamenica and Gentzkow (2011), studies strategic disclosure of information when the sender controls the information environment (called *signal*) and the receiver controls the choice of action to be taken. As a review by Kamenica (2019) suggests, this literature has provided many extensions of the original model of Kamenica and Gentzkow (2011) with interesting qualitative insights. However, full characterization of the optimal signal is generally difficult even in the original model. There has been little progress on this front, and it has been limited to a small number of special cases.[1]

We contribute to this literature by studying a special case of the original model that has received little attention – a Bayesian persuasion model in which both the sender and the receiver have *state-dependent preferred actions*. We characterize a qualitative property of the optimal signal called *state-pooling structure*, which describes pools of states that cannot be discerned from one another by the optimal signal. Specifically, we ask how the structure of state-dependent preference misalignment affects the state-pooling structure of the optimal signal.

To illustrate the main point of this paper, we present an example of a politician (receiver, he) and his advisor (sender, she). They both wish to implement some level of government spending $a \in \mathbb{R}$ that is adapted to the current economic situation captured by GDP per capita y, which takes one of three possible values: 1, 2, or 3. However, they each have a different vision of optimal spending as a function of GDP per capita. The advisor's payoff is $u_A(a, y) = -(a - \omega_A(y))^2$ and the politician's payoff is $u_P(a, y) = -(a - \omega_P(y))^2$, where $\omega_A(y)$ and $\omega_P(y)$ represent the preferred spending of the advisor and the politician in state y, respectively. The advisor designs an investigation (a signal) that can inform the politician about the realization of GDP per capita. She does that strategically to influence the spending choice of the politician. We are interested in how the structure of this

[1]We return to this point in the discussion of related literature in Section 2.

signal depends on the form of misalignment between the advisor's and politician's preferences captured by ω_A and ω_P, respectively.

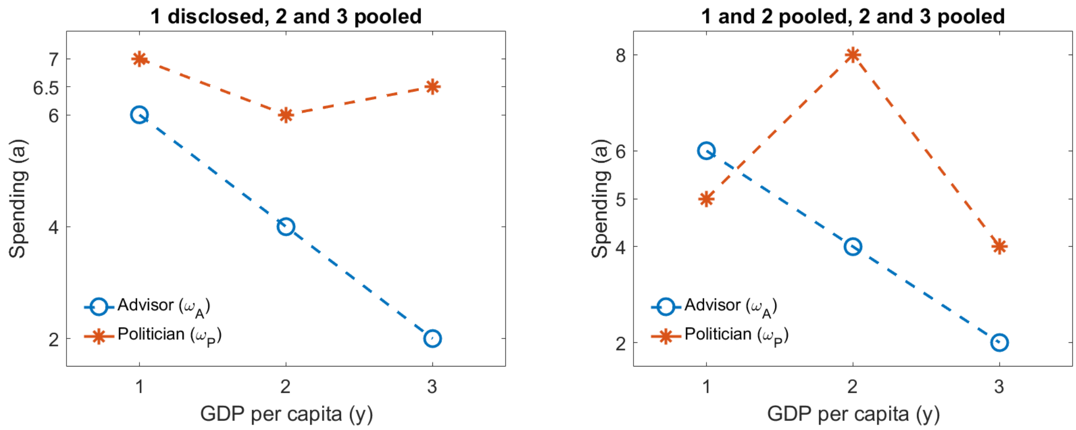

Figure 1: The form of disagreement between the advisor (ω_A) and the politician (ω_P) matters for the structure of the optimal signal:
left plot: the advisor fully reveals state 1 and pools states 2 and 3 together;
right plot: the advisor pools states 1 and 2 together and states 2 and 3 together (note: we consider only three levels of GDP per capita; the lines are drawn only for clarity of the picture)

Figure 1 illustrates how the form of disagreement between the advisor's and politician's preferred spending influences the structure of the optimal signal.[2] In the case presented *in the left plot*, the advisor's optimal signal fully reveals whether the state of the economy is low or not, i.e., one of the two outcomes of her investigation fully reveals the low state and the other leaves the politician uncertain about the high and middle states – we say they are pooled together. Intuitively, both the advisor and the politician want the highest spending in the low state, so their goals are aligned in this state and the advisor wants to reveal it perfectly. However, they disagree about whether the spending should be higher in the middle or high state, so the advisor wants to attenuate this disagreement by pooling these two states together. In the case presented *in the right plot*, the advisor's optimal signal reveals whether the economy is above or below average, i.e., one of the two outcomes of her investi-

[2]The structures of the optimal signals for the two cases considered in Figure 1 are derived using results from Section 6.

gation pools the low and middle states, while the other pools the middle and high states. Intuitively, the advisor and the politician disagree about whether the spending should be higher in the low or middle state, so the advisor wants to attenuate this disagreement by pooling these two states together. However, they both agree that the spending should be higher in the middle state than in the high state, but the politician prefers a greater spending difference between these two states than the advisor. Therefore, the advisor wants to moderate the politician's actions by pooling these two states together.

In Section 3, we describe our model. We use the Bayesian persuasion framework of Kamenica and Gentzkow (2011) with one-dimensional finite state space – the sender's preferred action. Both the sender and the receiver have quadratic loss functions with bliss points depending on the state of the world. The structure of misalignment is captured by function ρ mapping the state of the world (the sender's preferred action) to the receiver's preferred action. The case of linear ρ with slope 1 corresponds to the benchmark of perfect alignment.[3] We do not impose any requirements on this function and we analyze the role of its shape for the qualitative structure of the optimal signal in terms of state pooling.

In Section 4, we present general results on the pooling structure of the optimal signal. The patterns of pooling are driven by the sender's trade-off between (i) the informativeness of the signal, which leads to better adaptation of the action to the state of the world in states of alignment, and (ii) the revelation of the realized mismatch of the sender's and receiver's preferred actions, which drives the action of the receiver away from the sender's preferred action. First, we show that the sender generically benefits from revealing some information. The only cases in which non-disclosure is optimal are when ρ is linear with a slope sufficiently different from 1. Second, we demonstrate that the optimal signal does not induce an interior belief (except in cases of non-disclosure).

[3] A state-independent intercept does not affect the choice of the signal because it is a "sunk cost" for the sender.

In Section 5, we propose a simple graph procedure to characterize the optimal structure of state pooling for a given ρ. This procedure consists of an analysis of ρ on pairs of states and a test of pooling of more than two states. The crucial element of this procedure is the slope of ρ between pairs of states, which plays the role of an index of misalignment – if it is too high (disagreement about magnitude) or lower than zero (disagreement about order), then it indicates space for pooling; otherwise, it indicates space for separation.

In Section 6, we provide a full characterization of the state-pooling structure in the case of three states of the world. The state-pooling structure is completely pinned down by the shape of ρ except for the case in which ρ has a slope sufficiently different from 1 for each of the three pairs of states. In that case, the choice of a particular state-pooling structure depends both on the shape of ρ and the prior.

2 Related literature

First, we relate our work to the *Bayesian persuasion* literature. The most relevant results from the seminal paper by Kamenica and Gentzkow (2011) are (i) conditions for full disclosure or non-disclosure in the general form and (ii) comparative statics of more aligned preferences. Regarding point (i), we go beyond these two "corner" cases for the optimal signal, similarly as in the recent studies of Arieli et al. (2020) and Kolotilin and Wolitzky (2020). We discuss the connection of our work to Kolotilin and Wolitzky (2020) in more detail later in this section. Regarding point (ii), we perform a different exercise with preference misalignment: we fix the preferences and analyze how the structure of preference misalignment is related to the structure of state pooling of the optimal signal.

The methodological progress in Bayesian persuasion on the front of providing a general characterization of the structure of the optimal signal has been scarce. First, with two or three states of the world, concavification provides an insightful graphical method of solving the sender's problem (Kamenica and Gentzkow, 2011). Second, when the sender's utility depends only on the expected state, the "Rothschild-Stiglitz approach" (Gentzkow and Kamenica, 2016) and linear programming methods (Kolotilin, 2018; Dworczak and Martini, 2019) have been used to solve these problems. However, we are interested in situations with the *sender's state-dependent preferred* action and the role of the structure of preference misalignment, where these methods do not deliver immediate answers. We propose a new concavification-based approach of characterizing the state-pooling structure of the optimal signal.

The closest paper to ours is Kolotilin and Wolitzky (2020). However, we differ along several directions, and our paper can be viewed as complementary to theirs. First, their sender prefers higher actions independently of the state, but experiences state-dependent loss from mismatching the preferred action. In contrast, our sender has state-dependent preferred actions, but her loss from mismatching the preferred action is state-independent. Second, their receiver prefers higher actions in higher states; we do not impose this assumption. Third, they provide sufficient (and "almost necessary") conditions for special patterns of "assortative" disclosure. However, they do not provide a procedure for finding the pooling structure of the optimal signal explicitly, and they avoid characterization of more complicated patterns. In contrast, we work in a more specialized quadratic setting and do not restrict ourselves to characterization of specific (pairwise) pooling structures. Instead, we propose a general procedure for finding the pooling structure. Finally, the mechanisms driving the results in the two papers are different: in Kolotilin and Wolitzky (2020), the information does not have value for the sender alone, so state pooling emerges from pure persuasion concerns, while state pooling in our model is driven by the interplay of the sender's incentives to disclose the state and to hide misalignment.

Two other related papers in Bayesian persuasion literature are Alonso and Camara (2016) and Galperti (2019). Similar to our paper, both rely on the concavification technique to obtain insights regarding the optimal signal. Alonso and Camara (2016) consider the standard Bayesian persuasion model, but assume that the sender and the receiver have heterogeneous prior beliefs. While the sender in Alonso and Camara (2016) uses the variation of the difference between the sender's and receiver's prior beliefs across the states of the world to design the optimal disclosure, our sender uses the variation in the misalignment of the sender's and receiver's bliss points across the states of the world.[4] Galperti (2019) considers the standard Bayesian persuasion model in which the sender and the receiver have a special type of heterogeneous prior beliefs: the receiver attaches zero probability to some states that are perceived with positive probability by the sender. While we restrict attention to a sender with state-dependent bliss actions and study the general patterns of state pooling, Galperti (2019) makes weaker assumptions about preferences and focuses on patterns of pooling of the states that have a priori zero probability for the receiver.

Second, the results of our study are connected to the *literature on persuasion games*, in which the sender chooses how to disclose her private verifiable information regarding the state of the world. Milgrom (1981) and Milgrom and Roberts (1986) analyze the conventional model of a persuasion game and establish the result on "unraveling" of the sender's private information leading to full disclosure. Dye (1985) and Shin (1994) study state pooling in a similar game but with (second-order) uncertainty of the receiver about whether the sender actually has some private information or not. Seidmann and Winter (1997) analyze a persuasion game in which the sender has state-dependent preferred actions, and they demonstrate that the "unraveling" result still holds. The combination of these two features – second-order uncertainty and state-dependent preferred actions – has been studied in a small number of recent papers. The closest paper to ours is Hummel et al. (2018), in which unraveling does not occur due to the presence of the receiver's second-order uncertainty. In the Bayesian persuasion model that we study, the sender's disclosure mechanism serves a similar role to the one in Hummel et al. (2018): the sender moderates the

[4] They demonstrate that, under some mild conditions on the sender's and receiver's preferences, the sender generically chooses at least partial disclosure over non-disclosure. Similarly, in our model, the non-disclosure conditions are stringent.

receiver's actions via pooling of the states for which the sender's bliss-point line is sufficiently flat relative to that of the receiver.

Finally, Miura (2018) studies how pooling equilibria can be characterized based on a procedure that uses a *masquerade graph* introduced in Hagenbach et al. (2014). In his procedure, a pool of states is formed by the types of the sender who are mutually interested in masquerading, i.e., being perceived by the receiver as some other type in the pool. In spirit, this resembles the procedure for discovery of the state-pooling structure we introduce: a masquerade edge between two nodes (types) in Miura's graph procedure plays a similar role as an edge between two nodes (states) in our graph procedure – it captures a motive for manipulative non-disclosure.

3 Model

We consider the standard Bayesian persuasion framework: a sender (S, she) designs and commits to an information structure (a Blackwell experiment) about an unknown state of the world $\omega \in \Omega$ to influence the action $a \in A$ of a receiver (R, he). The state space is finite, $\Omega \subset \mathbb{R}$, $|\Omega| = n$, and the action space is continuous, $A = \mathbb{R}$. The sender and the receiver have a common prior $p_0 \in \Delta(\Omega)$. They have the following preferences:

$$u_S = -(a - \omega)^2,$$
$$u_R = -(a - \rho(\omega))^2,$$

where $\rho : \Omega \to \mathbb{R}$ is arbitrary. Hence, state ω represents the preferred action of the sender and $\rho(\omega)$ the preferred action of the receiver.[5]

[5]This model can be seen as a reduced form of a model in which the state of the world is two-dimensional, $y = (\omega_S, \omega_R)$, and the sender can design the experiment only about the dimension that is relevant for her, ω_S. The receiver then forms expectations about his relevant dimension, ω_R, using a common prior $p_0 \in \Delta(\Omega^2)$, so $\rho(\omega_S) = \mathrm{E}_{p_0}[\omega_R|\omega_S]$. This formulation maps better to the example with a politician and his advisor presented in the Introduction.

As is standard, the sender can be seen equivalently as choosing a Bayes-plausible distribution over posteriors, which we refer to as *signal*: $\pi \in \Delta(\Delta(\Omega))$ such that

$$\sum_{p \in supp(\pi)} \pi(p)p(\omega) = p_0(\omega) \ \forall \omega \in \Omega.^{6} \tag{1}$$

The timing is as follows: the sender chooses a signal π, a posterior belief p is drawn according to π, and the receiver takes an action a given the belief p. The solution concept is subgame perfect equilibrium. Going backwards, the receiver's optimal action given a posterior belief p is $a(p) = \mathrm{E}_p[\rho(\omega)]$. Hence, the game reduces to the following problem of the sender:

$$\max_{\pi \in \Delta(\Delta(\Omega))} -\mathrm{E}_\pi\left[\mathrm{E}_p\left[(\mathrm{E}_p[\rho(\omega)] - \omega)^2\right]\right] \text{ s.t. } \sum_{p \in supp(\pi)} \pi(p)p = p_0, \tag{2}$$

where $\mathrm{E}_\pi[\cdot]$ is the expectation over posteriors with respect to π and $\mathrm{E}_p[\cdot]$ is the expectation over states with respect to p.

4 General results about the optimal signal

In this section, we present general results about the optimal signal, and combine them in the next section to construct the procedure that allows us to discover which states are "pooled" together in the optimal signal.

To better understand how the sender chooses the signal, we start by inspecting the trade-off she faces. We can rewrite the objective function from her problem (2) as

$$\mathrm{var}_\pi(\mathrm{E}_p[\omega]) - \mathrm{E}_\pi\left[(\mathrm{E}_p[\omega - \rho(\omega)])^2\right]. \tag{3}$$

The first term captures the benefit of a more informative (in the sense of Blackwell) π – ideally, she would like to reveal all states perfectly.[7] The second term captures

[6] Kamenica and Gentzkow (2011) show that there exists an optimal π such that $|supp(\pi)| \leq \min\{|\Omega|, |A|\}$. Hence, we restrict our search for the optimal signal only to signals satisfying $|supp(\pi)| \leq n$.

[7] To illustrate this point, imagine an interior prior p_0, a signal π^1 with only interior beliefs, and a signal π^2 similar to π^1, but with more extreme beliefs: $p_k^2 = p_k^1 + \varepsilon(p_k^1 - p_0) \ \forall k$, for some small enough $\varepsilon > 0$. Then, $\mathrm{var}_{\pi^2}(\mathrm{E}_p[\omega]) = (1+\varepsilon)^2 \mathrm{var}_{\pi^1}(\mathrm{E}_p[\omega]) > \mathrm{var}_{\pi^1}(\mathrm{E}_p[\omega])$.

the "cost" of revealed misalignment – ideally, she would like to "pool" some states to hide the largest misalignment. Hence, the sender prefers to reveal the most information so that the action is well adapted to the state. However, since she does not control the action directly, she wants to exploit the form of misalignment captured by ρ to manipulate the action of the receiver.

We can notice that the intercept of ρ does not play a role for the optimal signal. Formally, consider any function ρ and take $\rho' = b + \rho$ for some arbitrary constant $b \in \mathbb{R}$. The sender's objective function

$$\mathrm{var}_\pi \left(\mathrm{E}_p \left[\omega\right]\right) - \mathrm{E}_\pi \left[\left(\mathrm{E}_p \left[\omega - \rho'(\omega)\right]\right)^2\right] \tag{4}$$

can be rewritten in the form

$$\mathrm{var}_\pi \left(\mathrm{E}_p \left[\omega\right]\right) - \mathrm{E}_\pi \left[\left(\mathrm{E}_p \left[\omega - \rho(\omega)\right]\right)^2\right] - 2b\mathrm{E}_{p_0} \left[\omega - \rho(\omega)\right] + b^2. \tag{5}$$

The last two terms in (5) do not depend on π, so the optimal signals under ρ and ρ' coincide. Hence, a state-independent bias b (no matter how large) does not affect the optimal signal.[8] Intuitively, the state-independent bias acts as a sunk cost for the sender. She cannot hide it by any manipulation of the signal because it is perfectly known ex ante.

It follows from the irrelevance of the intercept of ρ that what matters for the optimal signal is the overall shape of ρ, not agreement in particular states. In particular, perfect agreement between the sender and the receiver about the preferred action in a state of the world does not suffice for disclosure of that state. For example, consider two states $\omega_1 < \omega_2$, $\rho\left(\omega_1\right) = \omega_1$, $\rho\left(\omega_2\right) = 2\omega_1 - \omega_2$. Even though the sender and the receiver perfectly agree about the preferred action in ω_1, they substantially disagree in ω_2. It will be evident from the results in this section that full disclosure of the "perfect-agreement state" ω_1 is not optimal. Intuitively, due to the Bayesian consistency constraint, full disclosure of ω_1 would limit the opportunity to moderate the substantial disagreement in ω_2.[9]

[8] We can contrast this feature with cheap talk (Crawford and Sobel, 1982) in which the value of b matters for the informativeness of the equilibrium communication.

[9] In fact, Proposition 1 will imply that it is optimal not to disclose anything in this example.

4.1 Characterization of non-disclosure

In this subsection, we characterize the situation in which the sender does not benefit from revealing any information to the receiver.

Proposition 1. *The sender never (i.e., for any prior) benefits from providing any information if and only if* ρ *is linear with the slope from* $(-\infty, 0] \cup [2, +\infty)$.

Proof. The proof is in Appendix A. It identifies the conditions for concavity of the expected utility of the sender as a function of the induced posterior by the principal-minor test of the Hessian matrix of this function. □

Surprisingly, it is relatively easy to introduce some information revelation in our setting: it is sufficient to have a nonlinearity in ρ. The intuition for this generic taste for information revelation is that information has high value for the sender who wants to match the state of the world. The cases of optimal non-disclosure identified in Proposition 1 are intuitive too: (i) *misalignment in order*, i.e., when the sender and the receiver disagree about the order of the bliss actions (slope of ρ negative) or (ii) *misalignment in magnitude*, i.e., when they agree about the order, but the receiver overreacts relative to the sender (slope of ρ greater than two).

The non-disclosure characterized in Proposition 1 is never uniquely optimal for $n \geq 3$. To resolve such cases of indifference, we make the following assumption.

Assumption 1. Under indifference, the sender chooses not to disclose the states.

This assumption can be justified by the sender's interest in saving effort on communication when it is not needed. Technically, it greatly simplifies the analysis. Substantively, it leads us to identify the least informative signal in the indifference set of the sender. In Appendix B, we analyze the structure of our problem that gives rise to the cases of indifference, and discuss the role of Assumption 1 as opposed to other selection criteria.

4.2 Full disclosure

In the next proposition, we provide a sufficient condition for full disclosure of the state of the world.

Proposition 2. *If ρ is linear with a slope in $[0,2]$, full revelation of the state is always optimal (i.e., for any prior).*

Proof. The proof is in Appendix A. It mostly follows from the proof of Proposition 1. □

For general n, Proposition 2 provides only a sufficient condition for full disclosure, but for $n = 2$ we can provide a full characterization. This special case is a cornerstone of our analysis of the case with general n.

Lemma 1. *For $n = 2$, the sender strictly prefers full revelation if and only if the slope of ρ is in $(0,2)$. The sender is indifferent between any feasible signals if and only if the slope of ρ is either zero or two. The sender strictly prefers no revelation if and only if the slope of ρ is in $(-\infty,0) \cup (2,\infty)$.*

Proof. The proof is in Appendix A. □

4.3 "Extremization" – non-existence of an interior posterior

After analyzing the conditions for extreme signals (non-disclosure and full disclosure), we look at more structured signals. The following proposition provides the key result enabling that analysis.

Proposition 3 (Extremization)**.** *Suppose non-disclosure is not optimal. Then, it is never optimal to induce an interior posterior.*

Proof. The proof is in Appendix A. It is constructed by contradiction with the optimality of the signal, based on an improvement by splitting one of its posteriors.

We call this result "extremization" because it leads us from the interior of the simplex to its extreme (boundary) subsimplexes. □

We can apply Proposition 3 iteratively to eliminate the areas of posteriors that will not appear in the optimal signal. This sharpens the idea about the structure of the optimal signal, which is our main interest, and simplifies the search for it. We use this idea in the next section.

5 State-pooling structure of the optimal signal

In this section, we go beyond the extreme cases of full disclosure and non-disclosure and study how preference misalignment, captured by ρ, affects a qualitative property of the optimal signal that we call state pooling. We define the state-pooling structure of a signal and present an illustrative procedure for its discovery that builds on the general results from Section 4.

5.1 Definitions

Definition 1. We say that states $\omega_{k_1}, \ldots, \omega_{k_m}$, for some $k_1, \ldots, k_m \in \{1, \ldots, n\}$, are *pooled* together (or form a *pool of states*) under signal π if the set $M = \{\omega_{k_1}, \ldots, \omega_{k_m}\}$ satisfies

$$\exists p \in supp(\pi) : supp(p) = M \;\&\; \forall p' \in supp(\pi) \text{ s.t. } p' \neq p : M \nsubseteq supp(p'), \tag{6}$$

where $supp(\cdot)$ denotes support.[10] The set of all pools of states that signal π induces is called the *state-pooling structure* of signal π.

[10]In intuitive terms, $\omega_{k_1}, \ldots, \omega_{k_m}$ are pooled together under signal π if π reveals whether the event $\{\omega_{k_1}, \ldots, \omega_{k_m}\}$ occurred.

The state-pooling structure of a signal can be captured *graphically* by representing each state of the world by a node and each pool by highlighting the corresponding set of nodes; an example is presented in Figure 2.

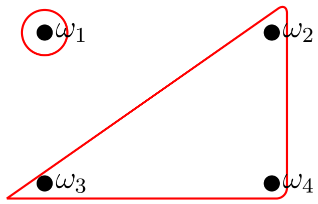

Figure 2: Example of a graphical representation of the state-pooling structure when $n = 4$ and the signal induces posteriors supported on $\{\omega_1\}$ and $\{\omega_2, \omega_3, \omega_4\}$

In the next subsection, we propose a procedure that aims to find the state-pooling structure of the optimal signal for a given form of preference misalignment captured by ρ. This procedure can easily be represented graphically; its desired output is a graphical representation of the state-pooling structure of the type depicted in Figure 2, i.e., nodes representing states and highlighted pools. However, the proposed procedure may not identify the state-pooling structure of the optimal signal completely in some cases, but may offer only candidates for optimal pools. Nevertheless, we can often identify which of the candidate pools are certainly a part of the optimal state-pooling structure. Hence, we introduce two types of highlighting in the procedure – *dashed* (highlighting candidate pools) and *full* (highlighting pools certainly belonging to the optimal state-pooling structure). Naturally, highlighting in full is superior to highlighting in dashed because it expresses certainty.

An important working component of the graphical procedure is the *edges* between pairs of nodes – they represent a pooling tendency of the corresponding states. We will see that this pooling tendency is driven by the slope of ρ between pairs of corresponding states; we denote the *slope of ρ between states ω_i and ω_j* by

$$s_{ij} = \frac{\rho(\omega_j) - \rho(\omega_i)}{\omega_j - \omega_i}. \tag{7}$$

This object represents an index of misalignment between the receiver (the numerator) and the sender (the denominator).[11]

A subroutine of our procedure relates to the well-known problem from computer science called the *clique problem*. Thus, we borrow a few notions from graph theory.

Definition 2. Let $G = (V, E)$ be an undirected graph (with V denoting the set of nodes and E denoting the set of edges). We call a subset of nodes $C \subseteq V$ *clique* if the subgraph of G induced by C is complete (i.e., the nodes in C are fully connected). A *clique* C is called *maximal* if there does not exist another clique strictly above C (in the sense of inclusion).

The version of the clique problem that we are interested in is finding all maximal cliques in an undirected graph. Systematic inspection of all subsets of nodes or the *Bron–Kerbosch algorithm* can be used to solve this problem.

5.2 Procedure for discovery of the state-pooling structure of the optimal signal

We present a procedure that inspects the form of misalignment function ρ and reflects its implications for the state-pooling structure of the optimal signal on a graph. The output are pools highlighted in full (which are certainly present in the state-pooling structure of the optimal signal) and candidate pools highlighted in dashed (which may be present in the state-pooling structure of the optimal signal). We present an example of the output of this procedure at the end of this subsection and a step-by-step illustration of the procedure leading to this output in Appendix C.

[11] A similar object plays an important role for the pooling structure (of types) in Hummel et al. (2018).

Procedure for discovery of the state-pooling structure of the optimal signal:

Input: Set of states Ω ($|\Omega| = n$) and preference-misalignment function $\rho : \Omega \to \mathbb{R}$.

1. Create a fully connected graph on n nodes where node i corresponds to state ω_i.
2. Eliminate all edges ij such that the slope of ρ on $\omega_i < \omega_j$, s_{ij}, is in $(0, 2)$.
3. Highlight in full each isolated node (i.e., a node with no edges leading to any other node) as a singleton pool.
4. Among the remaining (i.e., non-isolated) nodes, list all maximal cliques.
5. For each maximal clique C:

   for k from $|C|$ to 2:

   for all subsets $M \subseteq C$ such that $|M| = k$:

   - If M was ever inspected before, do nothing and continue iteration.
   - If M is a subset of a highlighted set of nodes, do nothing and continue iteration.
   - Otherwise, apply the non-disclosure test to the inspected pool M: Is ρ linear with slope in $(-\infty, 0] \cup [2, \infty)$ on the states corresponding to the nodes in M?
     - If yes, highlight pool M in dashed on output and continue iteration.
     - If no, denote M as inspected and continue iteration.
6. If any node belongs only to one highlighted pool (in dashed), highlight the corresponding pool in full (if not already highlighted in full).

An example of the output produced by this procedure appears in the right panel of Figure 3; an example of function ρ leading to this output is depicted in the left panel.[12] State 1 is isolated because the sender and the receiver agree on its position

[12] A step-by-step illustration of the procedure leading to this output appears in Appendix C.

relative to other states both in order and in magnitude, so there is no reason for the sender to leverage this state for manipulation of beliefs. States 2, 3, and 4 are pooled together (they pass the non-disclosure test) because the sender tries to moderate the action of the receiver, who would overreact in these states (disagreement about magnitude). States 3 and 5 may be pooled together (disagreement about order) and 4 and 5 may also be pooled together (disagreement about order), but states 3, 4, and 5 are not pooled together even though they form a maximal clique (because they do not pass the non-disclosure test) – the sender prefers to exploit some variation in this collection of nodes. Hence, the optimal signal will induce posterior $p_1 = \delta_1$ and posterior p_2 supported on 2, 3, and 4. Moreover, it will induce at least one of the posteriors p_3 or p_4 supported on 3 and 5 or 4 and 5, respectively.

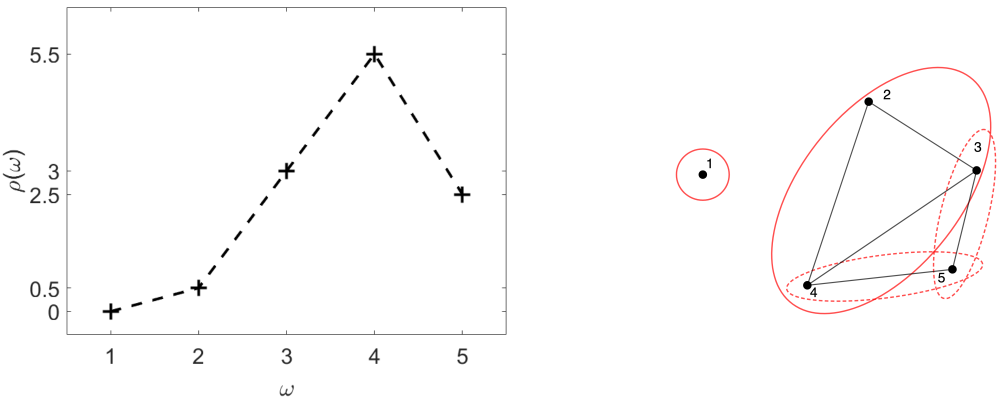

Figure 3: Output of the graph procedure (right panel) for function ρ on the left panel: 1 is isolated; 2, 3, and 4 are pooled together (they pass the non-disclosure test); 3 and 5 may be pooled together; 4 and 5 may be pooled together; 3, 4, and 5 are not pooled together (they do not pass the non-disclosure test)

5.3 Discussion of the procedure

The idea underlying our proposed procedure is the iterative application of Proposition 1 and Proposition 3, which we call a *top-down approach*. Starting from the full

$(n-1)$-dimensional simplex,[13] we can check whether non-disclosure is optimal using Proposition 1. If it is optimal, the sender chooses a completely uninformative signal. If it is not, Proposition 3 suggests that the optimal signal will induce posteriors on the boundary of the $(n-1)$-dimensional simplex. Hence, we focus on each of the $(n-2)$-dimensional boundary simplexes and apply the same test. Specifically, by restricting the sender's expected utility (as a function of the posterior) on a particular $(n-2)$-dimensional simplex, we use Proposition 1 to check if non-disclosure is optimal there:

- If it is optimal, then the sender cannot benefit from splitting the pool of states corresponding to the vertices of the inspected $(n-2)$-dimensional simplex. However, the sender might not want to choose this pool of states at all, so this pool of states constitutes only a candidate pool for the optimal signal.[14]
- If it is not optimal, then by Proposition 3 we eliminate all interior points from the inspected $(n-2)$-dimensional simplex and restrict our focus to its $(n-3)$-dimensional boundary simplexes; for each of them, we repeat the same steps.

Along the path from the full $(n-1)$-dimensional simplex to lower-dimensional simplexes due to elimination of "interior" posteriors outlined in the second bullet point, we move closer to the trivial case of 1-dimensional simplexes where we apply Lemma 1.

Our procedure relies on this top-down approach in Step 5. However, compared to the top-down approach, the procedure starts with a simplification of the problem by identifying the only relevant subsets of nodes for this inspection – the maximal cliques (Steps 2 and 4). This step is justified by the fact that the necessary condition for optimality of non-disclosure on a simplex is optimality of non-disclosure on its

[13]We start from the $(n-1)$-dimensional simplex because $p_n = 1 - p_1 - \cdots - p_{n-1}$.

[14]Here, we also use Assumption 1. This simplifies the analysis because we do not need to keep track of all equivalent splits.

boundary simplexes, which follows easily from Proposition 1. Hence, if we have a given collection of nodes with some pair of nodes in it that is not pooled, this whole collection of nodes cannot form a pool.

In Steps 3 and 6 of the procedure, we exploit Bayesian consistency (and the interior prior). In particular, the structure of the graph obtained after Step 2 is informative about the state-pooling structure by itself: any isolated node represents a state that is fully disclosed. In Step 5, we can identify only candidates for optimal pools, but, in Step 6, Bayesian consistency can help us to determine which of them will be certainly a part of the optimal pooling structure.

Note that we have not mentioned the prior in our identification of the optimal pooling structure. This prior-independence of our procedure relies on a feature of the quadratic setting: constant convexity/concavity structure in all points. However, even in the quadratic setting, the pooling structure of the optimal signal itself is not always prior-independent. This feature imposes a limit on how far we can go with our simple prior-independent procedure in identifying the full pooling structure of the optimal signal. In some cases, we also need to incorporate the prior into our analysis at the end of the procedure (see Section 6 for examples).

6 Characterization of the state-pooling structure for $n = 3$

In this section, we use the above procedure to characterize the state-pooling structure of the optimal signal in the simplest interesting case of three states (the case of two states is trivial and is fully characterized in Lemma 1). We describe the state-pooling structure for all possible cases of the form of ρ, which we capture through s_{12}, s_{23}, and s_{13}. For clarity of exposition, we divide the cases into five classes (i)-(v) based on the features of the resulting state-pooling structure and the role of the prior. Class (i) corresponds to full disclosure, class (ii) corresponds to signals that fully disclose one of the states, classes (iii) and (iv) correspond to signals that reveal some information without fully revealing any of the states, and class (v) corresponds to non-disclosure. Within a given class, we use letters to distinguish between particular state-pooling structures.

Proposition 4. *Assume that there are three states of the world, $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Depending on the form of ρ, as pinned down by s_{12}, s_{23}, and s_{13}, the state-pooling structure of the optimal signal is as follows:*

	s_{12}	s_{23}	s_{13}	state-pooling structure
i	$\in (0,2)$	$\in (0,2)$	$\in (0,2)$	$\{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$
ii.a	$\in (0,2)$	$\notin (0,2)$	$\in (0,2)$	$\{\{\omega_1\}, \{\omega_2, \omega_3\}\}$
ii.b	$\notin (0,2)$	$\in (0,2)$	$\in (0,2)$	$\{\{\omega_3\}, \{\omega_1, \omega_2\}\}$
iii.a	$\notin (0,2)$	$\in (0,2)$	$\notin (0,2)$	$\{\{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}\}$
iii.b	$\in (0,2)$	$\notin (0,2)$	$\notin (0,2)$	$\{\{\omega_2, \omega_3\}, \{\omega_1, \omega_3\}\}$
iii.c	$\notin (0,2)$	$\notin (0,2)$	$\in (0,2)$	$\{\{\omega_1, \omega_2\}, \{\omega_2, \omega_3\}\}$
iv[15]	$\notin (0,2)$	$\notin (0,2)$	$\notin (0,2)$	depending on s_{12}, s_{23}, s_{13}, and prior, either (iii.a), (iii.b), or (iii.c) pooling
v	$s_{12} = s_{23} = s_{13} = s \notin (0,2)$			$\{\{\omega_1, \omega_2, \omega_3\}\}$

Proof. The proof is in Appendix A. □

The observed state-pooling structures emerge from the interaction of the two main forces that drive the sender's choice. On the one hand, the sender wants to disclose the states so that the induced receiver's actions vary sufficiently with the state of the world. On the other hand, she wants to pool the states together to dampen

[15] $s_{12} = s_{23} = s_{13} = s \notin (0,2)$ corresponds to non-disclosure, so we exclude this combination from case (iv) and denote it as a separate case (v). See Appendix A for details on the choice from (iii.a), (iii.b), and (iii.c).

that variation if there is a severe misalignment in either order or magnitude in some pairs of states. The slope of ρ for states ω_i and ω_j, s_{ij} ($i, j \in \{1, 2, 3\}$, $i \neq j$), serves as an index that can capture the misalignment in either order or magnitude in that pair of states.

In case (i), there is no severe preference misalignment in either pair of states, so the sender fully discloses each state. In case (ii.a), s_{23} captures a severe preference misalignment in the pair of states ω_2, ω_3, so the sender pools these states together to conceal the misalignment but reveals state ω_1 to maximize the informativeness of the signal. In case (iii.a), s_{12} and s_{13} capture a severe preference misalignment in two pairs of states, so the sender pools the respective pairs together but still reveals some information: $\{\{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}\}$. In case (iv), there is a misalignment in each of the three pairs of states and the optimal state-pooling structure is sensitive to the prior and to the relation between the slopes of ρ.

A notable feature of the state-pooling structure of the optimal signal under $n = 3$ is that the sender never chooses to fully disclose the middle state of the world ω_2 and pool ω_1 and ω_3 together. For that to be the case, it would need to hold $s_{13} \notin (0, 2)$, $s_{12} \in (0, 2)$, and $s_{23} \in (0, 2)$, which cannot happen.[16] The intuition is that full disclosure of ω_2 and pooling of ω_1 and ω_3 is not in line with the sender's preference for maximizing the variance of the induced posterior beliefs. A potentially better way to leverage state ω_2 is to form two pools $\{\omega_2, \omega_1\}$ and $\{\omega_2, \omega_3\}$ because it can induce relatively more variation in the receiver's actions.

7 Conclusion

We consider a Bayesian persuasion model in which both the sender and the receiver have state-dependent preferred actions. We specialize to a quadratic-utility setting to simplify the otherwise nontrivial problem of characterizing the optimal signal. In

[16] Note that $s_{13} = \frac{\rho(\omega_3)-\rho(\omega_1)}{\omega_3-\omega_1} = \frac{1}{(\omega_3-\omega_2)+(\omega_2-\omega_1)} \left(s_{23}(\omega_3 - \omega_2) + s_{12}(\omega_2 - \omega_1)\right)$ and $(0, 2)$ is a convex set.

this framework, we make the trade-off that drives the sender's choice of the signal transparent: on the one hand, the sender wants to reveal information to adapt the action to the state of the world; on the other hand, she wants to hide information to conceal the misalignment between her and the receiver.

We focus on characterization of the state-pooling structure of the optimal signal. In particular, we link the form of misalignment between the sender and the receiver in their preferred (state-dependent) actions to the state-pooling structure of the sender's optimal signal. To achieve this goal, we propose an illustrative graphical procedure for finding the sets of states that are pooled together in the supports of posteriors of the optimal signal.

Our model naturally suits the analysis of influence in political economy. The sender's and receiver's (state-dependent) single-peaked preferences over the continuous action space are consistent with ideology-based preferences over a continuous set of policy alternatives. That set could represent potential allocations of a resource such as the amount of budget spending on a public good. Thus, our framework can capture an arbitrary form of ideological disagreement between a lobbyist and a policymaker regarding the preferred state-dependent policy and yield predictions about the structure of the lobbyist's chosen information disclosure.

Our analysis motivates a number of directions for further research. First, further investigation and economic interpretation of particular state-pooling patterns that emerge when there are more than three states of the world might be of interest. Second, more progress could be made on analyzing state-pooling patterns that may emerge under loss functions of a more general form.

References

Alonso, R. and Camara, O. (2016). Bayesian persuasion with heterogeneous priors. *Journal of Economic Theory*, 165:672–706.

Arieli, I., Babichenko, Y., Smorodinsky, R., and Yamashita, T. (2020). Optimal persuasion via bi-pooling. In *Proceedings of the 21st ACM Conference on Economics and Computation*, EC '20, page 641, New York, NY, USA. Association for Computing Machinery.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451.

Dworczak, P. and Martini, G. (2019). The simple economics of optimal persuasion. *Journal of Political Economy*, 127(5):1993–2048.

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

Galperti, S. (2019). Persuasion: The art of changing worldviews. *American Economic Review*, 109(3):996–1031.

Gentzkow, M. and Kamenica, E. (2016). A Rothschild-Stiglitz approach to Bayesian persuasion. *American Economic Review*, 106(5):597–601.

Hagenbach, J., Koessler, F., and Perez-Richet, E. (2014). Certifiable pre-play communication: Full disclosure. *Econometrica*, 82(3):1093–1131.

Hummel, P., Morgan, J., and Stocken, P. C. (2018). A model of voluntary managerial disclosure. Unpublished manuscript.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272.

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

Kolotilin, A. (2018). Optimal information disclosure: A linear programming approach. *Theoretical Economics*, 13(2):607–635.

Kolotilin, A. and Wolitzky, A. (2020). Assortative information disclosure. UNSW Economics Working Paper 2020-08.

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

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

Miura, S. (2018). Prudence in persuasion. Unpublished manuscript.

Seidmann, D. J. and Winter, E. (1997). Strategic information transmission with verifiable messages. *Econometrica: Journal of the Econometric Society*, pages 163–169.

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

Appendix

A Technical details and proofs

A.1 The structure of the sender's problem

We are interested in the solution of the sender's problem

$$\max_{\pi \in \Delta(\Delta(\Omega))} -\mathrm{E}_{\pi}\left[\mathrm{E}_{p}\left[(\mathrm{E}_{p}\left[\rho(\omega)\right] - \omega)^{2}\right]\right] \text{ s.t. } \sum_{p \in supp(\pi)} \pi(p)p = p_{0}. \tag{8}$$

We can rewrite the objective function as

$$-\mathrm{E}_{\pi}\left[\mathrm{E}_{p}\left[\rho(\omega)\right]^{2} - 2\mathrm{E}_{p}\left[\rho(\omega)\right]\mathrm{E}_{p}\left[\omega\right] + \mathrm{E}_{p}\left[\omega^{2}\right]\right]. \tag{9}$$

Using the Bayesian consistency condition $\sum_{p} \pi(p)p = p_{0}$, we can see that the last term becomes

$$-\mathrm{E}_{p_{0}}\left[\omega^{2}\right]. \tag{10}$$

Therefore, the solution to the problem above is the same as the solution to the problem

$$\max_{\pi \in \Delta(\Delta(\Omega))} \mathrm{E}_{\pi}\left[\mathrm{E}_{p}\left[\rho(\omega)\right]\left(2\mathrm{E}_{p}\left[\omega\right] - \mathrm{E}_{p}\left[\rho(\omega)\right]\right)\right] \text{ s.t. } \sum_{p \in supp(\pi)} \pi(p)p = p_{0}. \tag{11}$$

A general approach to solving this problem is concavification of the function

$$g(p) = \mathrm{E}_{p}\left[\rho(\omega)\right]\left(2\mathrm{E}_{p}\left[\omega\right] - \mathrm{E}_{p}\left[\rho(\omega)\right]\right). \tag{12}$$

We use the parametrization $g(p) = g(p_1, p_2, \ldots, p_{n-1})$, where $p_n = 1 - p_1 - \cdots - p_{n-1}$. We collect the free variables in the vector

$$\bar{p} = (p_1, \ldots, p_{n-1})'.$$

We also denote

$$\bar{\rho} = (\rho(\omega_1) - \rho(\omega_n), \ldots, \rho(\omega_{n-1}) - \rho(\omega_n))',$$
$$\bar{\omega} = (\omega_1 - \omega_n, \ldots, \omega_{n-1} - \omega_n)'.$$

With this notation, we can write

$$g(\bar{p}) = \bar{p}' \underbrace{[2\bar{\rho}\bar{\omega}' - \bar{\rho}\bar{\rho}']}_{G} \bar{p} + [2\omega_n\bar{\rho}' - \rho_n\bar{\rho}' + 2\rho_n\bar{\omega}' - \rho_n\bar{\rho}']\bar{p} + 2\rho_n\omega_n - \rho_n^2. \tag{13}$$

Hence, the curvature of g is driven by matrix G because the Hessian matrix is

$$H = G + G'.^{17} \tag{15}$$

The ij element $(i, j \in \{1, \ldots, n-1\})$ of H is

$$H_{ij} = \frac{\partial^2 g(p)}{\partial p_i \partial p_j} = 2\{[\rho(\omega_i) - \rho(\omega_n)](\omega_j - \omega_n) - [\rho(\omega_i) - \rho(\omega_n)][\rho(\omega_j) - \rho(\omega_n)] \\ + [\rho(\omega_j) - \rho(\omega_n)](\omega_i - \omega_n)\}. \tag{16}$$

This special structure of the problem implies that general submatrices of order 3 (for $n \geq 4$) of the Hessian matrix H have zero determinants.[18] Hence, by the Laplace expansion of determinants, all submatrices of order $k \geq 3$ have zero determinants. We can deduce from this observation, using the fact that the determinant rank of a matrix is equal to the column/row rank of the matrix,[19] that H has at most two non-zero eigenvalues. Therefore, there are at least $n-3$ orthogonal directions (in space $\mathbb{R}^{n-1} \ni \bar{p}$) that span the space along which g is linear, and at most two orthogonal directions that span the space (orthogonal to the space spanned by the linear directions) on which g has a less trivial shape.

[17] We can also rewrite g as a linear-quadratic form

$$g(\bar{p}) = \frac{1}{2}\bar{p}' H \bar{p} + [2\omega_n\bar{\rho}' - \rho_n\bar{\rho}' + 2\rho_n\bar{\omega}' - \rho_n\bar{\rho}']\bar{p} + 2\rho_n\omega_n - \rho_n^2. \tag{15}$$

[18] Proof is available upon request. It is basically just tedious algebra.

[19] The *determinant rank* of H is the size k of the largest $k \times k$ submatrix with a non-zero determinant. The *column/row rank* of H is the dimension of the space spanned by the columns/rows of H. It is straightforward to show that these ranks are equal.

A.2 Proofs

Proof of Proposition 1. The sender does not benefit from providing any information if and only if g is concave.[20] g is concave if and only if its Hessian matrix is negative semidefinite, which can be checked with the test on its principal minors.

Suppose $n \geq 3$ (the case $n = 2$ is covered separately in Lemma 1). Let Δ_k be a principal minor of order k of the Hessian matrix of g. Since $\Delta_k = 0$ for $k \geq 3$ (see the discussion above), a necessary and sufficient condition for g to be concave is $\Delta_1 \leq 0$ and $\Delta_2 \geq 0$ for all Δ_1, Δ_2.

Let Δ_1^i be the first-order principal minor obtained from row (column) i:

$$\Delta_1^i = 2\left(\rho\left(\omega_i\right) - \rho\left(\omega_n\right)\right)\left(2\left(\omega_i - \omega_n\right) - \left(\rho\left(\omega_i\right) - \rho\left(\omega_n\right)\right)\right). \tag{17}$$

Let Δ_2^{ij} be the second-order principal minor obtained from rows (columns) i and j:

$$\Delta_2^{ij} = -4[(\rho(\omega_i) - \rho(\omega_j))(\omega_j - \omega_n) - (\rho(\omega_j) - \rho(\omega_n))(\omega_i - \omega_j)]^2. \tag{18}$$

We can see that $\Delta_2^{ij} \leq 0$. Hence, g is concave or convex only if $\Delta_2 = 0$ for all Δ_2. This condition yields a system of $\frac{(n-1)(n-2)}{2}$ equations

$$\Delta_2^{ij} = 0,\ i, j \in \{1, ..., n-1\},\ i \neq j. \tag{19}$$

Under the natural assumption that $\omega_1 < \cdots < \omega_n$ (which is without loss of generality), we obtain from $\Delta_2^{ij} = 0$

$$\frac{\rho\left(\omega_j\right) - \rho\left(\omega_i\right)}{\omega_j - \omega_i} = \frac{\rho\left(\omega_n\right) - \rho\left(\omega_j\right)}{\omega_n - \omega_j} \tag{20}$$

or, equivalently,

$$\frac{\rho\left(\omega_j\right) - \rho\left(\omega_i\right)}{\omega_j - \omega_i} = \frac{\rho\left(\omega_n\right) - \rho\left(\omega_i\right)}{\omega_n - \omega_i}. \tag{21}$$

[20] The "if" part follows directly from the definition of concavity. The "only if" part would also follow directly from the definition of concavity if the sender did not benefit from providing any information for every prior. But if the sender does not benefit from providing any information only in one prior, because g is a linear-quadratic form, this property extends to all priors.

Therefore, the system of equations (19) gives rise to $\frac{(n-1)(n-2)}{2}$ slope equality conditions. From (20) and (21), we have

$$
\begin{aligned}
j = n-1, i = n-2: \frac{\rho(\omega_n) - \rho(\omega_{n-1})}{\omega_n - \omega_{n-1}} &= \frac{\rho(\omega_{n-1}) - \rho(\omega_{n-2})}{\omega_{n-1} - \omega_{n-2}} = \frac{\rho(\omega_n) - \rho(\omega_{n-2})}{\omega_n - \omega_{n-2}}, \\
j = n-2, i = n-3: \frac{\rho(\omega_n) - \rho(\omega_{n-2})}{\omega_n - \omega_{n-2}} &= \frac{\rho(\omega_{n-2}) - \rho(\omega_{n-3})}{\omega_{n-2} - \omega_{n-3}} = \frac{\rho(\omega_n) - \rho(\omega_{n-3})}{\omega_n - \omega_{n-3}}, \\
&\vdots \\
j = 2, i = 1: \frac{\rho(\omega_n) - \rho(\omega_2)}{\omega_n - \omega_2} &= \frac{\rho(\omega_2) - \rho(\omega_1)}{\omega_2 - \omega_1} = \frac{\rho(\omega_n) - \rho(\omega_1)}{\omega_n - \omega_1}.
\end{aligned}
$$

Hence, system (19) is equivalent to a linearity of ρ:

$$s := \frac{\rho(\omega_2) - \rho(\omega_1)}{\omega_2 - \omega_1} = \frac{\rho(\omega_3) - \rho(\omega_2)}{\omega_3 - \omega_2} = \cdots = \frac{\rho(\omega_n) - \rho(\omega_{n-1})}{\omega_n - \omega_{n-1}}. \tag{22}$$

Finally, given that $\Delta_2 = 0$ for all Δ_2 holds, one can establish whether g is concave or convex based on the sign of Δ_1. Inspecting the sign of (17) yields:

$$\Delta_1^i \geq 0 \iff (\rho(\omega_n) - \rho(\omega_i) \geq 0) \wedge \frac{\rho(\omega_n) - \rho(\omega_i)}{\omega_n - \omega_i} \leq 2 \iff 0 \leq s \leq 2. \tag{23}$$

The complement identifies the concavity slopes (including the borderline slopes $s \in \{0, 2\}$). □

Proof of Proposition 2. This proposition is basically proven in the proof of Proposition 1, using the fact that g is convex if and only if $\Delta_1 \geq 0$ and $\Delta_2 \geq 0$ for all Δ_1, Δ_2. The only difference is that the convexity of g is only sufficient for optimality of full disclosure, but is not necessary (we can provide an example of optimal full disclosure with non-convex g). □

Proof of Lemma 1. For $n = 2$, g is a quadratic function, so its second derivative completely characterizes its curvature, which completely characterizes the type of optimal signals. In particular, let $\omega_1 < \omega_2$. Then,

$$\frac{\partial^2 g(p_1)}{\partial p_1^2} = 2(\rho(\omega_1) - \rho(\omega_2))(2(\omega_1 - \omega_2) - (\rho(\omega_1) - \rho(\omega_2))), \tag{24}$$

which is strictly positive if and only if the slope of ρ is in $(0, 2)$ (strict convexity and full disclosure), strictly negative if and only if the slope of ρ is in $(-\infty, 0) \cup (2, \infty)$

(strict concavity and non-disclosure), and zero if and only if the slope of ρ is either zero or two (linearity and indifference). □

Proof of Proposition 3. Non-disclosure is optimal if and only if g is concave. Hence, if non-disclosure is not optimal, g is not concave. Therefore, g has to have a direction along which it is strictly convex.[21]

Suppose (toward contradiction) that it is optimal to induce an interior posterior, i.e., there exists a posterior p in the support of the optimal signal π such that $p(\omega) > 0 \ \forall \omega$. Then, we can split p along a strictly convex direction to q_1 and q_2, i.e., there exists some $\lambda \in (0,1)$ such that $p = \lambda q_1 + (1-\lambda) q_2$. Then, π' formed from π by replacing p by q_1 with probability $\lambda \pi(p)$ and q_2 with probability $(1-\lambda)\pi(p)$ is Bayes-plausible and it induces a strict improvement for the sender because, from strict convexity of g along the direction determined by q_1 and q_2,

$$\mathrm{E}_{\pi'}\left[g(p)\right] - \mathrm{E}_{\pi}\left[g(p)\right] = \pi(p)(\lambda g(q_1) + (1-\lambda) g(q_2) - g(p)) > 0. \tag{25}$$

This is a contradiction with optimality of π. □

Proof of Proposition 4. We derive the state-pooling structure for the form of ρ for each case presented in the table of Proposition 4 using the graph procedure presented in Section 5.2.

Case (i). Since $s_{12}, s_{23}, s_{13} \in (0,2)$, Step 2 of the procedure eliminates all edges, so each node is highlighted in full in Step 3. Thus immediately after Step 3, the procedure yields the state-pooling structure of the optimal signal $\{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}\}$.

Case (ii.a). Since $s_{12}, s_{13} \in (0,2)$ and $s_{23} \notin (0,2)$, after Step 2 of the procedure, node 1 is isolated (thus, it is highlighted in full in Step 3) and there is an edge left between nodes 2 and 3. Since the pool $\{2,3\}$ is a maximal clique (Step 4) and ρ is obviously linear with slope from $(-\infty, 0] \cup [2, \infty)$ on states ω_2 and ω_3, this pool is highlighted in dashed in Step 5. Finally, it is highlighted in full in Step 6 because

[21] This is independent of the position because g is a linear-quadratic form.

nodes 2 and 3 belong only to this pool. Therefore, the state-pooling structure of the optimal signal is $\{\{\omega_1\}, \{\omega_2, \omega_3\}\}$.

Case (ii.b). Analogous to case (ii.a).

Case (iii.a). Since $s_{12}, s_{13} \notin (0, 2)$ and $s_{23} \in (0, 2)$, after Step 2 of the procedure, there are two edges left: one between nodes 1 and 2 and one between nodes 1 and 3. Since both pools $\{1, 2\}$ and $\{1, 3\}$ are maximal cliques (Step 4) and ρ is obviously linear with slope from $(-\infty, 0] \cup [2, \infty)$ on states ω_1, ω_2 and ω_1, ω_3, respectively, these pools are highlighted in dashed in Step 5. Finally, they are highlighted in full in Step 6 because node 2 belongs only to pool $\{1, 2\}$ and node 3 belongs only to pool $\{1, 3\}$. Therefore, the state-pooling structure of the optimal signal is $\{\{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}\}$.

Case (iii.b). Analogous to case (iii.a).

Case (iii.c). Analogous to case (iii.a).

Case (iv). We assume that $s_{12} = s_{23} = s_{13} = s \notin (0, 2)$ does not hold (this case is covered by case (v)). Thus, the graph procedure yields the candidate pools $\{\omega_1, \omega_2\}$, $\{\omega_2, \omega_3\}$, and $\{\omega_1, \omega_3\}$ (corresponding to the pools of nodes highlighted in dashed in the graph). To determine the optimal state-pooling structure given the set of candidate pools is non-trivial.

Denote the n-th directional derivative of a function $f : \mathbb{R}^2 \to \mathbb{R}$ along a direction (a, b) by $D^n_{(a,b)} f$. Denote $p_1 := \Pr(\omega_1)$ and $p_2 := \Pr(\omega_2)$. From the proof of Proposition 1, the nonlinearity in ρ implies that there exists a direction (a, b) along which $g(p_1, p_2)$ (defined in (12)) is strictly convex. The set of all such directions is pinned down by the condition

$$D^2_{(a,b)} g(p) > 0, \tag{26}$$

which rewrites as (assuming $s_{13} \neq 0$ and $s_{23} \neq 0$; see below for the discussion of these cases)

$$\begin{aligned}
a^2 \left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_1-\omega_3\right)-\left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\right]+ \\
b^2 \left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_2-\omega_3\right)-\left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\right]+ \\
ab\tfrac{\rho(\omega_2)-\rho(\omega_3)}{\rho(\omega_1)-\rho(\omega_3)} \left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_1-\omega_3\right)-\left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\right]+ \\
ab\tfrac{\rho(\omega_1)-\rho(\omega_3)}{\rho(\omega_2)-\rho(\omega_3)} \left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_2-\omega_3\right)-\left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\right]>0.
\end{aligned} \tag{27}$$

Next, $s_{13} \notin (0,2) \wedge s_{23} \notin (0,2)$ implies[22]

$$\begin{cases}
\left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_1-\omega_3\right)-\left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\right] & \leq 0, \\
\left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_2-\omega_3\right)-\left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\right] & \leq 0.
\end{cases} \tag{28}$$

We can see from (27) and (28) that if (a,b) is a direction along which g is strictly convex, both a and b have to be non-zero. Thus, we can normalize the direction (a,b) to $(\frac{a}{b},1)$ and denote $x := \frac{a}{b}$. Hence, the set of directions along which g is strictly convex is characterized by

$$\begin{aligned}
x^2 \left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_1-\omega_3\right)-\left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\right]+ \\
\left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_2-\omega_3\right)-\left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\right]+ \\
x\tfrac{\rho(\omega_2)-\rho(\omega_3)}{\rho(\omega_1)-\rho(\omega_3)} \left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_1-\omega_3\right)-\left(\rho\left(\omega_1\right)-\rho\left(\omega_3\right)\right)\right]+ \\
x\tfrac{\rho(\omega_1)-\rho(\omega_3)}{\rho(\omega_2)-\rho(\omega_3)} \left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\left[2\left(\omega_2-\omega_3\right)-\left(\rho\left(\omega_2\right)-\rho\left(\omega_3\right)\right)\right]>0.
\end{aligned} \tag{29}$$

Inspecting (29) given (28), one observes that the first two terms in (29) are non-positive. Therefore, the sum of the last two terms must necessarily be strictly positive for any direction along which g is strictly convex. Further, if the third term is strictly negative, the fourth term is non-positive and vice versa. So, if either of the last two terms is strictly negative, their sum is also strictly negative. Equivalently, if their sum is non-negative, they both have to be non-negative. Moreover, if their sum is strictly positive, they cannot both be zero. But if any one of the last two terms in (29) is strictly positive, then by (28)

$$x\frac{\rho(\omega_1)-\rho(\omega_3)}{\rho(\omega_2)-\rho(\omega_3)} < 0. \tag{30}$$

[22] At least one of these terms is non-zero due to the assumption that $s_{12} = s_{23} = s_{13} = s \notin (0,2)$ does not hold.

To summarize, if $(x, 1)$ is a direction along which g is strictly convex, then

$$\begin{cases} x > 0 & \text{if } \frac{\rho(\omega_1)-\rho(\omega_3)}{\rho(\omega_2)-\rho(\omega_3)} < 0 \ (\iff \frac{s_{13}}{s_{23}} < 0), \\ x < 0 & \text{if } \frac{\rho(\omega_1)-\rho(\omega_3)}{\rho(\omega_2)-\rho(\omega_3)} > 0 \ (\iff \frac{s_{13}}{s_{23}} > 0). \end{cases} \tag{31}$$

By similar arguments, if $s_{13} = 0$,[23] the necessary condition for $(x, 1)$ being the direction along which g is strictly convex is

$$\begin{cases} x > 0 & \text{if } s_{23} > 0, \\ x < 0 & \text{if } s_{23} < 0 \end{cases} \tag{32}$$

and if $s_{23} = 0$, the necessary condition for $(x, 1)$ being the direction along which g is strictly convex is

$$\begin{cases} x > 0 & \text{if } s_{13} > 0, \\ x < 0 & \text{if } s_{13} < 0. \end{cases} \tag{33}$$

Given some interior prior, the sender splits it along a direction along which g is strictly convex and induces posteriors that lie on two edges of the simplex. We can distinguish the following cases:

1. If $\frac{s_{13}}{s_{23}} < 0$ or $s_{13} = 0 \wedge s_{23} > 0$ or $s_{23} = 0 \wedge s_{13} > 0$, then $x > 0$. Hence, the optimal split is either of the form $(q_1, 0, 1-q_1)$, $(1-q_2, q_2, 0)$ (pooling case (iii.a)) or of the form $(q_1, 1-q_1, 0)$, $(0, q_2, 1-q_2)$ (pooling case (iii.c)) depending on the prior.

2. If $\frac{s_{13}}{s_{23}} > 0$ or $s_{13} = 0 \wedge s_{23} < 0$ or $s_{23} = 0 \wedge s_{13} < 0$, then $x < 0$. In this case, we need to distinguish further:

   (a) If the optimal split goes along the direction $(-1, 1)$, it is of the form $(q_1, 0, 1-q_1)$, $(0, q_2, 1-q_2)$ (pooling case (iii.b)).

   (b) If the optimal split goes along direction $(x, 1)$ with $x < -1$, it is either of the form $(q_1, 0, 1-q_1)$, $(0, q_2, 1-q_2)$ (pooling case (iii.b)) or of the form $(q_1, 1-q_1, 0)$, $(0, q_2, 1-q_2)$ (pooling case (iii.c)) depending on the prior.

[23] Notice that s_{13} and s_{23} cannot be simultaneously zero by assumption, because this would lead to case (v).

(c) If the optimal split goes along direction $(x, 1)$ with $x > -1$, it is either of the form $(q_1, 0, 1-q_1)$, $(0, q_2, 1-q_2)$ (pooling case (iii.b)) or of the form $(q_1, 0, 1-q_1)$, $(q_2, 1-q_2, 0)$ (pooling case (iii.a)) depending on the prior.

Case (v). Proposition 1 applies and under Assumption 1 yields non-disclosure. □

B Comment on Assumption 1

The structure of function g (see (12)) uncovered in Section A.1 implies that for $n \geq 4$, there always exists a direction along which g is linear. Therefore, even when g is concave and non-disclosure is optimal, it is never uniquely optimal for $n \geq 4$. In particular, the sender is indifferent between sticking to the prior and splitting it to some posteriors from the space determined by the linear directions of g (and the prior), possibly all the way to the boundaries of the original simplex. Moreover, if g is concave, it is also concave on the boundary simplexes and we can repeat the same argument, proceeding downward in dimensions. For $n = 3$, by Proposition 1, g is concave only if it is linear in one direction. Hence, even for $n = 3$, non-disclosure is not uniquely optimal and the sender is indifferent between choosing a non-informative signal (keeping the belief at the prior) and splitting the prior into posteriors along the linear direction, all the way to the edges of the simplex. Therefore, pairwise signals (i.e., signals leading to posteriors supported on at most two states) are also always optimal.[24]

In the main text, we impose Assumption 1, which resolves indifference in favor of non-disclosure of states. It is a natural assumption that can be justified by the sender not wasting resources (time and energy) on communication when it is not needed (although the cost of communication is not featured explicitly in our model). This selection criterion simplifies the analysis. First, it enables us to avoid imposing some ad hoc assumptions about the selection of specific partial disclosure patterns

[24]This result is reminiscent of the result of Kolotilin and Wolitzky (2020) that there is no loss of generality from focusing on pairwise signals in their setup.

from the indifference set. Second, a different natural assumption might be that the sender resolves her indifference in favor of splitting. However, this assumption would require us to impose some additional ad hoc assumptions about the selection of specific directions along which to split (for higher n) in order to deliver concrete predictions. Moreover, such a resolution of indifference would be very sensitive to the prior (even in terms of the predicted pooling structure), so we would need to keep track of the specific directions of indifference, which would render the analysis much more cumbersome.[25]

C Demonstration of the procedure for discovery of the state-pooling structure of the optimal signal

We demonstrate the application of the procedure for discovery of the state-pooling structure of the optimal signal (presented in Section 5) to the example introduced in Figure 3 (for convenience, we reproduce it in Figure 4 in this section). This demonstration is accompanied by Figure 5. *Red color* in Figure 5 represents highlighting as defined in Section 5 – final pools in full and candidate pools in dashed. *Green color* denotes cliques chosen for application of the non-disclosure test (Step 5 of the procedure).

[25]To illustrate the dependence on the prior, for $n = 3$ under linear ρ (which is sufficient for global concavity or convexity), the direction of linearity is $(-\frac{\omega_3-\omega_2}{\omega_3-\omega_1}, 1)'$. Since the first component is strictly between 0 and -1, we can see that, while the non-disclosure is also optimal, the state-pooling structure (defined in Section 5) of the optimal informative signal can be either $\{\{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}\}$ or $\{\{\omega_1, \omega_3\}, \{\omega_1, \omega_2\}\}$, depending on the prior.

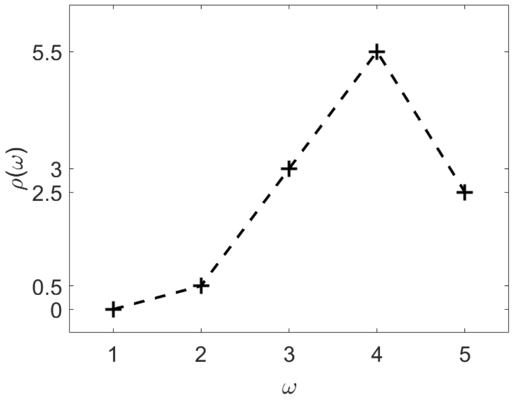

Figure 4: Preference misalignment function ρ considered for the demonstration of the graph procedure

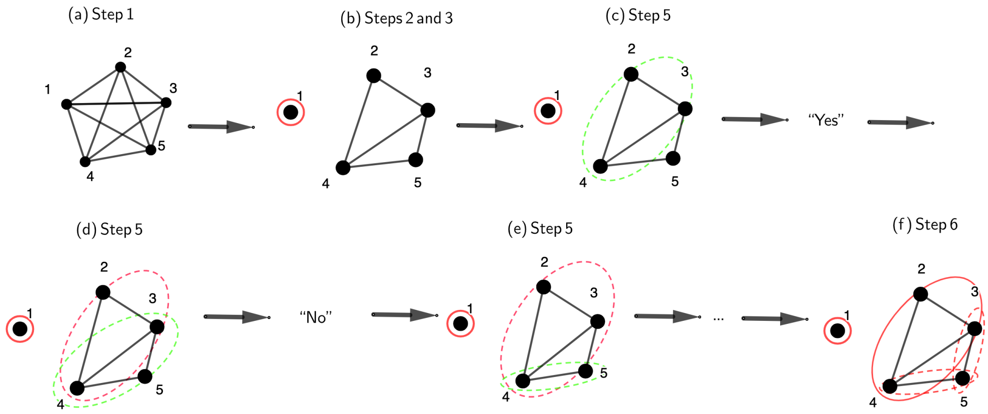

Figure 5: Illustration of the execution of the procedure, applied to the input from Figure 4; the output is in (f); red color represents highlighting as defined in Section 5 – final pools in full and candidate pools in dashed; green color denotes cliques chosen for application of the non-disclosure test

The inputs to the procedure are the values of ω and $\rho(\omega)$ from Figure 4. From formula (7), we obtain the values of all s_{ij}: $s_{12} = 0.5$, $s_{13} = 1.5$, $s_{14} = \frac{5.5}{3}$, $s_{15} = \frac{2.5}{4}$, $s_{23} = 2.5$, $s_{24} = 2.5$, $s_{25} = \frac{2}{3}$, $s_{34} = 2.5$, $s_{35} = -\frac{1}{4}$, $s_{45} = -3$.

In (a) in Figure 5, we start with a fully connected graph on five nodes ($n = 5$) corresponding to states $1, 2, ..., 5$.

In (b) in Figure 5, we observe the same graph after the application of Steps 2 and 3 of the procedure. We removed all edges ij such that $s_{ij} \in (0, 2)$. As a result, node 1 became isolated, so we highlighted it in full. Hence, we can leave out node 1 from further analysis and focus on nodes 2, 3, 4, and 5.

In (c) in Figure 5, we proceed to Steps 4 and 5 of the procedure. It is easily seen that there are two maximal cliques: one formed by nodes 2, 3, and 4 and one formed by nodes 3, 4, and 5. First, we inspect the maximal clique formed by 2, 3, and 4 (highlighted in green) and we apply the non-disclosure test. The non-disclosure condition holds, so we highlight the maximal clique $\{2, 3, 4\}$ in dashed (as illustrated in (d)). Hence, we do not need to consider any more of its subsets in Step 5 and we can move our focus to the other maximal clique.

In (d) in Figure 5, we inspect the maximal clique formed by nodes 3, 4, and 5 (highlighted in green). The non-disclosure condition does not hold, so we denote the maximal clique $\{3, 4, 5\}$ as inspected and proceed to consider its subsets of cardinality 2.

In (e) in Figure 5, we first consider the clique formed by nodes 4 and 5. As the non-disclosure condition is satisfied, we highlight this clique in dashed. Proceeding with the iteration, we test clique $\{3, 5\}$. Again, the non-disclosure condition is satisfied, so we highlight it in dashed. Finally, clique $\{3, 4\}$ is a subset of the highlighted set $\{2, 3, 4\}$, so we do not test it.

In (f) in Figure 5, we proceed to Step 6 of the procedure: as node 2 belongs to only one highlighted clique, $\{2, 3, 4\}$, we highlight that clique in full. The output of the procedure is depicted in (f) in Figure 5: the singleton pool $\{1\}$ and pool $\{2, 3, 4\}$ highlighted in full and pools $\{3, 5\}$ and $\{4, 5\}$ highlighted in dashed. Hence, the posteriors induced by the optimal signal certainly include a posterior supported on states $\omega_2 = 2, \omega_3 = 3, \omega_4 = 4$ and the posterior δ_{ω_1}. Moreover, the optimal signal will induce at least one posterior supported on $\omega_3 = 3, \omega_5 = 5$ or $\omega_4 = 4, \omega_5 = 5$.

Abstrakt

V této práci se zabýváme modelem Bayesiánskeho přesvědčování s konečně mnoha stavy a kvadratickými ztrátovými funkcemi odesílatele a příjemce závisejícími na stavu. Nesouhlas mezi odesílatelem a příjemcem ohledně optimální akce může mít libovolný tvar. Tento model umožňuje zaměřit se na relativně neprozkoumaný kompromis mezi informativností signálu a utajením nesouhlasu ohledně optimální akce. Konkrétně se zaměřujeme na to, jak odesílatel sdružuje stavy v posteriorech optimálního signálu. Předkládáme ilustrativní grafovou proceduru, která vyžaduje na vstupu formu nesouhlasu ohledně optimální akce (jako funkci stavu) a na výstupu produkuje potenciální reprezentace struktury sdružování stavů při optimálním signálu. Naše analýza může přispět k porozumění situacím, kde odesílateli a příjemci záleží na různých aspektech společného objektu zájmu, např. komunikace mezi politickým poradcem, kterému záleží na stavu ekonomiky, a politikem, kterému záleží na politické situaci.

Klíčová slova: Bayesiánské přesvědčování, strategické sdružování stavů, tvar nesouladu preferencí, grafová procedura

Working Paper Series
ISSN 1211-3298
Registration No. (Ministry of Culture): E 19443

Individual researchers, as well as the on-line and printed versions of the CERGE-EI Working Papers (including their dissemination) were supported from institutional support RVO 67985998 from Economics Institute of the CAS, v. v. i.

Specific research support and/or other grants the researchers/publications benefited from are acknowledged at the beginning of the Paper.

Published by
Charles University, Center for Economic Research and Graduate Education (CERGE)
and
Economics Institute of the CAS, v. v. i. (EI)
CERGE-EI, Politických vězňů 7, 111 21 Prague 1, tel.: +420 224 005 153, Czech Republic.
Printed by CERGE-EI, Prague
Subscription: CERGE-EI homepage: http://www.cerge-ei.cz

Phone: + 420 224 005 153
Email: office@cerge-ei.cz
Web: http://www.cerge-ei.cz

Editor: Byeongju Jeong

The paper is available online at http://www.cerge-ei.cz/publications/working_papers/.

ISBN 978-80-7343-515-8 (Univerzita Karlova, Centrum pro ekonomický výzkum a doktorské studium)
ISBN 978-80-7344-610-9 (Národohospodářský ústav AV ČR, v. v. i.)