Constructive versus toxic argumentation in debates

Constructive Vs Toxic Argumentation in Debates

By Tymofiy Mylovanov and Andriy Zapechelnyuk*

Two debaters address an audience by sequentially choosing their information strategies. We compare the setting where the second mover reveals additional information (constructive argumentation) with the setting where the second mover obfuscates the first mover's information (toxic argumentation). We reframe both settings as constrained optimization of the first mover. We show that when the preferences are zero-sum or risk-neutral, constructive debates reveal the state, while toxic debates are completely uninformative. Moreover, constructive debates reveal the state under the assumption on preferences that capture autocratic regimes, whereas toxic debates are completely uninformative under the assumption on preferences that capture democratic regimes.
JEL: D82, D83, D72
Keywords: Information design, Bayesian persuasion, information structure, disclosure, obfuscation, garbling

I. Introduction

This paper attempts to formalize and compare two phenomena that are pervasive in communication conflicts, such as political debates. On the one hand, debating parties use constructive arguments, such as attestations of reputable experts, to inform the audience in a controlled way in order to achieve a desired effect. On the other hand, the parties sometimes deploy toxic arguments, such as scandalous or entertaining statements, to reduce the impact the opponent's arguments on the audience.

Toxic arguments are those that carry negative criticism, blame, and contempt. Negative criticism is deployed to point out that the opponent's arguments are inaccurate or unsound. Blame and contempt are used to target the opponent's personal flaws and to show that

* *Date:* May 9, 2023

Mylovanov: University of Pittsburgh, Department of Economics, 4714 Posvar Hall, 230 South Bouquet Street, Pittsburgh, PA 15260, USA. *E-mail:* mylovanov@gmail.com.
Zapechelnyuk: School of Economics, University of Edinburgh, 31 Buccleuch Place, Edinburgh, EH8 9JT, UK. *E-mail:* azapech@gmail.com.

The authors are grateful for helpful comments and inspirational conversations to Alp Atakan, Yuriy Butusov, Selman Erol, David Jaeger, Keiichi Kawai, Stephan Lauermann, Elliot Lipnowski, Marco Mariotti, Ilia Murtazashvili, Peter Norman, Ronny Razin, Larry Samuelson, Daniel Seidmann, Nataliia Shapoval, Joel Sobel, Ina Taneva, Richard van Weelden, Alistair Wilson, Leeat Yariv, the participants of various seminars where this paper has been presented, and anonymous referees. Tymofiy Mylovanov acknowledges the support from the Office of Naval Research Multidisciplinary University Research Initiative (MURI) under award number N00014-17-1-2675 and from Kyiv School of Economics. Andy Zapechelnyuk acknowledges the support from the Economic and Social Research Council Grant ES/N01829X/1. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the supporting organizations.

the opponent is not worth listening to. The ultimate purpose of toxic argumentation is to cast a doubt on the validity and credibility of the opponent's arguments and reduce their informational content as perceived by the audience. Toxic argumentation stands in contrast with constructive argumentation that adds to rather than subtracts from the informational content of the opponent's arguments.

In this paper we use the concept of information disclosure (as in Bayesian persuasion) to model constructive argumentation, and the concept of information obfuscation, or garbling, to model toxic argumentation. We present a simple model that captures the distinction between the two kinds of argumentation and allows us to compare how they affect truth discovery. In our model, two debaters sequentially choose information disclosure strategies of an uncertain state of the world in order to influence the choice of a heterogeneous audience. We compare two cases: *sequential disclosure* and *sequential obfuscation*. In the case of sequential disclosure, the second mover reveals additional information about the state (referred to as constructive argumentation). In the case of sequential obfuscation, the second mover obfuscates the information revealed by the first mover (referred to as toxic argumentation).

We ask how and to what extent the nature of counterarguments by the second mover affects truth discovery by the audience. Answering this question will allow us to contribute to the policy discussion of whether debates should adhere to the principle of freedom of speech (thus potentially allowing toxic arguments) or whether they should be moderated, so that only constructive arguments are allowed.

At a glance, complementing one's argument with another informative argument should result in more information disclosure than obfuscating one's argument. But after a moment of reflection this should not be obvious. The first mover can adjust her behavior in anticipation of the opponent's counteraction. For example, she can strategically choose to disclose more information when expecting the opponent to obfuscate some of it. Furthermore, note that the case of sequential obfuscation can be equivalently represented literally, as two parties sequentially obfuscating an initially revealed state of the world. So, there is an intrinsic symmetry between sequential disclosure of an initially hidden state and sequential obfuscation of an initially revealed state. There is no difference when there is only one sender, and it is not obvious what difference it makes to the strategic interaction of two senders.

We begin by showing how the problems of sequential disclosure and sequential obfuscation can be simplified. After the simplification, the difference between the two problems becomes apparent. In both cases, the first mover solves a constrained optimization problem with the same objective but different constraints. In sequential disclosure, the first mover chooses among the outcomes that the second mover cannot improve upon by further disclosure. In contrast, in sequential obfuscation, the first mover chooses among the outcomes that the second mover cannot improve upon by obfuscation. This allows us to show that sequential obfuscation cannot make the audience better informed than sequential disclosure.

A substantial part of our analysis is devoted to determining the conditions under which *full disclosure* (i.e., the outcome that reveals the state to the audience) and *no disclosure* (i.e., the outcome that reveals nothing about the state) are obtained in equilibrium. An informal takeaway of this analysis is that sequential disclosure often leads to full disclosure, and sequential obfuscation often leads to no disclosure. The opposite is rare: For sequential

disclosure to result in no disclosure, it must be the case that no disclosure is the outcome that is Pareto dominant, that is, both parties prefer no disclosure to all other outcomes. Similarly, for sequential obfuscation to result in full disclosure, it must be the case that full disclosure is Pareto dominant.

In particular, we establish sufficient conditions for full disclosure (no disclosure) to be the equilibrium outcome of sequential disclosure (sequential obfuscation, respectively) in a few special cases which are notable in the literature and relevant for applications. Importantly, and perhaps surprisingly, these results are completely independent of the prior distribution of the state, so they apply regardless of how much the prior favors one party over the other. These results are as follows.

Suppose that both parties are risk neutral or have zero-sum preferences. Then we obtain a polarizing outcome: sequential disclosure leads to full disclosure, and at the same time sequential obfuscation leads to no disclosure. When the preferences are zero-sum, the result for the sequential disclosure game follows from the observation that each party can individually enforce full disclosure by unilaterally revealing the state. Consequently, the payoff of the full disclosure outcome becomes the "value" of the game. Risk neutral preferences are essentially zero-sum (up to scaling), because every unit of utility gained by one party translates to a constant number of units of utility lost by the other party. The symmetric statement holds for sequential obfuscation.

Then, we consider the case motivated by debates of political parties who are competing for citizens' support. Suppose that the distribution of the citizens' types in the audience has a log-concave probability density, and the parties have log-concave marginal utility functions. Such assumptions are common in economic applications and include several prominent special cases. A log-concave density exhibit nice properties, such as unimodality and hazard rate monotonicity. Many familiar probability density functions are log-concave (see Table 1 in Mark Bagnoli and Ted Bergstrom, 2005). A log-concave marginal utility has increasing Arrow-Pratt coefficient of risk aversion, so the more support a party gains, the more averse it is to gambling with this support. This assumption also captures the case in which the parties care more about obtaining the support of the citizens near the median of the population distribution (e.g., simple majority) and less about those at the extremes.

Consider the ratio of marginal utilities of the parties as a function of their support by the citizens. First, consider the case of a decreasing marginal utility ratio, which means that every utility unit gained by one party translates into an increasingly larger number of utility units lost by the other party. This is the case of democratic regimes where the minority party stands to gain more from increasing its support than the majority party stands to lose. When this is the case, we show that toxic debates (sequential obfuscation) reveal nothing about the state, and thus, they are informationally inferior to constructive debates (sequential disclosure). This provides a possible explanation why democratic regimes are often not very good in digging out the truth, and highlights the danger of negative criticism and contempt for truth discovery in political debates.

Second, consider the case of an increasing marginal utility ratio, which means that every utility unit lost by one party translates into an increasingly larger number of utility units gained by the other party. This is the case of authoritarian regimes where the majority party stands to gain more from squashing the minority opposition than the opposition stands to

lose. When this is the case, we show that constructive debates (sequential disclosure) always lead to full disclosure of the state, and thus, they are informationally superior to toxic debates (sequential obfuscation). Perhaps, this is why authoritarian regimes, which understand the threat of constructive debates in exposing the truth, are so keen to discredit or completely shut down the opposition.

Do our results imply that the society should regulate the freedom of speech to mitigate information obfuscation? This is a scary proposition in practice. Who will be the judge of what is considered toxic? The government? An appointed committee? The answer is outside of our formal model, but we hope that technological innovation driven by competition among social platforms will eventually take care of this. A recent (albeit fleeting) popularity of audio social networks such as Clubhouse or Audio Telegram provide an example. In audio social networks, discussions are moderated, often moderators allow one person to speak at a time, speakers are allowed to respond to accusations and comments, they face penalties (e.g., ban) for using toxic arguments, and interaction happens in real time with the audience present and focused on the speakers. The audience appears to be attracted to the platforms that are better moderated and have more informative discussions.

Related Literature. This paper contributes to the literature on competition in information design where senders commit to information disclosure protocols before learning the state of the world. Matthew Gentzkow and Emir Kamenica (2017*a*,*b*), Fei Li and Peter Norman (2018), and Dilip Ravindran and Zhihan Cui (2022) consider senders who simultaneously choose information structures. The peculiarity of simultaneous disclosure is that when more than one sender discloses the same bit of information, no sender can unilaterally prevent its disclosure. This leads to multiplicity of equilibria, in particular, full disclosure of the state is always an equilibrium. Raphael Boleslavsky and Christopher Cotton (2018) and Pak Hung Au and Keiichi Kawai (2020) restrict the senders to disclose different coordinates of a multidimensional state, thus preventing the overlap in the information disclosure. Fei Li and Peter Norman (2021) and Wenhao Wu (2022) consider sequential, rather than simultaneous disclosure, where sequential moves lead a unique equilibrium outcome.

There are several closely related papers to ours. Li and Norman (2021), Wu (2022), and Frederic Koessler, Marie Laclau, Jérôme Renault and Tristan Tomala (2022) study variations of sequential disclosure, and Itai Arieli, Yakov Babichenko and Fedor Sandomirskiy (2022) study sequential obfuscation. The settings of these papers are more general than ours. Li and Norman (2021), Wu (2022), and Arieli, Babichenko and Sandomirskiy (2022) feature multiple senders who move sequentially, whereas Koessler et al. (2022) have two senders who take turns in disclosing information over potentially infinite periods. The focus of these papers is on the general principles and methodology of the formulation of the problem and the derivation of equilibria using concavification and recursive derivation of constraints.

Our novelty relative to the above papers is twofold. First, we add a context-driven structure to the problem, and consequently we obtain results that are more meaningful and interpretable for applications. We focus on special, applicable cases where extreme outcomes (full disclosure and no disclosure) are obtained. Moreover, the latter results are robust as they are independent of the prior about the state. Second, to our knowledge, our paper is first to compare information disclosure and information obfuscation. When there is a single sender, obfuscation of an initially revealed state is strategically identical to disclosure

of an initially hidden state. In the information design literature with a single sender, the term *obfuscation* (garbling, confusion) appears synonymously to the term *disclosure* and is often used to emphasize the interpretation where the sender reduces information about an initially revealed state (e.g., Jimmy Chan, Seher Gupta, Fei Li and Yun Wang, 2019; Chris Edmond and Yang K Lu, 2021; Fei Li, Yangbo Song and Mofei Zhao, 2023). As seen from our paper, disclosure and obfuscation are not identical instruments when there is more than one sender.

In our paper we adopt a so-called linear information design approach. Linearity refers to the property that the payoffs depend on the posterior belief about the state only through the posterior mean. This approach received a lot of attention on the literature (Emir Kamenica and Matthew Gentzkow, 2011; Matthew Gentzkow and Emir Kamenica, 2016; Anton Kolotilin, Tymofiy Mylovanov, Andriy Zapechelnyuk and Ming Li, 2017; Anton Kolotilin, 2018; Anton Kolotilin and Andriy Zapechelnyuk, 2019; Piotr Dworczak and Giorgio Martini, 2019; Itai Arieli, Yakov Babichenko, Rann Smorodinsky and Takuro Yamashita, 2023; Andreas Kleiner, Benny Moldovanu and Philipp Strack, 2021). It has been used in many applications of information design, including media control (Scott Gehlbach and Konstantin Sonin, 2014; Boris Ginzburg, 2019; Arda Gitmez and Pooya Molavi, 2022; Anton Kolotilin, Tymofiy Mylovanov and Andriy Zapechelnyuk, 2022), clinical trials (Anton Kolotilin, 2015), voter persuasion (Ricardo Alonso and Odilon Câmara, 2016), transparency benchmarks (Darrell Duffie, Piotr Dworczak and Haoxiang Zhu, 2017), stress tests (Itay Goldstein and Yaron Leitner, 2018; Dmitry Orlov, Pavel Zryumov and Andrzej Skrzypach, 2022), online markets (Gleb Romanyuk and Alex Smolin, 2019), attention management (Elliot Lipnowski, Laurent Mathevet and Dong Wei, 2020; Alexander W Bloedel and Ilya Segal, 2020), quality certification (Andriy Zapechelnyuk, 2020; Benjamin Vatter, 2022), and healthcare congestion in epidemics (Ju Hu and Zhen Zhou, 2022).

Our paper is also related to the literature on competitive expertise and informational lobbying, where a policy maker or legislator consults two or more biased experts. A focal question in this literature is whether seeking advice of multiple experts can improve the information disclosure to the policy maker, and if so, whether full disclosure can be achieved. In Thomas W Gilligan and Keith Krehbiel (1989), Barton L Lipman and Duane J Seppi (1995), Vijay Krishna and John Morgan (2001*a*,*b*), Marco Battaglini (2002), Attila Ambrus and Satoru Takahashi (2008), Ming Li (2010), and Tymofiy Mylovanov and Andriy Zapechelnyuk (2013*a*,*b*) the experts know the state of the world, so consulting more than one expert has no informational benefit, but it can improve the incentives for information disclosure. Lipman and Seppi (1995) is worth a special mention, because in this model the experts can prove the correctness of certain type of messages, thus having a limited commitment power. In David Austen-Smith (1993), Hyun Song Shin (1998), Asher Wolinsky (2002), Marco Battaglini (2004), Gilat Levy and Ronny Razin (2007), and Attila Ambrus and Shih En Lu (2014), each expert's private information is partial, and consulting more that one expert can improve the informational content, whereas Li (2010) shows that more experts can result in less disclosure for strategic reasons.[1] The effects of the order in which experts present their arguments are investigated in Krishna and Morgan (2001*b*) and Elena D'Agostino and Daniel J Seidmann (2022), the collusion of the experts is explored in Andriy Zapechelnyuk (2013), and the experts' strategic decisions about how much information

[1]Li and Norman (2018) show a similar finding in a Bayesian persuasion setting.

to acquire are studied in Isabelle Brocas, Juan D Carrillo and Thomas R Palfrey (2012) and Faruk Gul and Wolfgang Pesendorfer (2012). Our paper contributes to this literature by addressing a complementary question about the difference between improvement and erosion of the informational content by an addition of an "expert".

Our main motivating story is that of a debate. The term *debate* refers to a decision procedure that formalizes rhetoric and argumentation, where informed but biased parties choose arguments, and an uninformed listener reaches a conclusion based on these arguments.[2] Jacob Glazer and Ariel Rubinstein (2001) study an abstract model where the state is a string of 0's and 1's, and the listener wants to know whether there are more 1's than 0's. They adopt a mechanism design approach: To elicit information from two informed parties, the listener designs a sequential communication protocol subject to a constraint on its complexity. Ran Spiegler (2006) studies a setting where two parties debate on two issues at the same time. He uses an axiomatic approach to derive a solution that describes how arguments should be selected and how winners should be chosen. Gilat Levy and Ronny Razin (2012) model a debate as an all-pay auction in which two parties bid for attention slots of a decision maker. Our paper adopts a more pragmatic interpretation of a debate as competition of two biased parties in information disclosure to citizens.

II. Model

A. Setup

Two parties are engaged in a debate on some issue relevant to the public, for example, whether some economic policy should be implemented, or whether an accusation against one of the parties is true and that party should face a political defeat. The two parties are called an *accuser* and a *respondent*, and labeled by A and R. The truth about the issue is summarized by a random unobserved state of the world $\theta \in \Theta = [0,1]$. The public consists of a continuum of citizens indexed by type $t \in T = [0,1]$. The type captures the heterogeneity of the citizens' attitudes towards the issue. The state θ and the type t are distributed independently, according to prior probability distribution functions F and G.

Each citizen needs to choose whether to support party A or party R. The citizens do not observe the state, but they receive information about it from the parties. Given a posterior expected value of the state, denoted by x, the utility of each citizen with type t is given by $x - t$ if the citizen decides to support party A, and it is equal to 0 if the citizen decides to support party R. In words, citizens with higher types t are more predisposed to support the respondent, and the higher their type, the higher the posterior value of the state should be to make them support the accuser instead.

The parties are expected utility maximizers. Their preferences are as follows. Let q_i be an expected fraction of citizens who support party $i = A, R$, so $q_A + q_R = 1$. Each party $i = A, R$ obtains the utility $u_i(q_i)$, which is twice continuously differentiable and strictly increasing in q_i. For example, the parties are interested in maximizing their public support

[2]In some public economics and political science literature, the term *debate* has a different meaning and refers to a pre-play cheap talk communication of asymmetrically informed legislators, e.g., David Austen-Smith (1990), Marco Ottaviani and Peter Sørensen (2001), and David Spector (2000).

on the debated issue, and they are risk averse, so their utilities $u_i(q_i)$ are concave in q_i. For another example, the parties are interested in reaching the support by the simple majority, so utility u_i smoothly approximates the step function that gives utility 0 when $q_i < 1/2$ and utility 1 when $q_i > 1/2$.

Let us describe the parties' strategies. Let M_i be a set of messages of party $i = A, R$. Suppose that the sets M_A and M_R are rich enough, so $\Theta \subseteq M_A$ and $\Theta \times M_A \subseteq M_R$. A strategy of party A is a mapping $\phi_A : \Theta \to \Delta(M_A)$ that associates with each state θ a conditional probability distribution $\phi_A(\cdot|\theta)$ over party A's messages in M_A. A strategy of party R is a mapping $\phi_R : \Theta \times M_A \to \Delta(M_R)$ that associates with each state θ and each message m_A of party A a conditional probability distribution $\phi_R(\cdot|\theta, m_A)$ over party R's messages in M_R. Importantly, because party R is the second mover who observes party A's strategy ϕ_A, throughout the paper it is understood that $\phi_R(\cdot|\theta, m_A)$ implicitly depends on ϕ_A.

The parties have full commitment to their strategies and have no discretion at the communication stages. The interpretation is that the parties make a lot of preparatory work for the debate: they invite experts, think up arguments and contingency responses, write scripts, and prepare supporting evidence. When the debate takes place, the parties are unable to deviate from what they have prepared, e.g., they cannot conjure new arguments or evidence they have not made ready in advance, and they cannot control what their experts are saying.

The timing is as follows. Parties A and R choose their strategies sequentially. Then state θ realizes. Then, message m_A is generated according to party A's strategy, after which message m_R is generated according to party R's strategy. The citizens observe the strategies of the parties and message m_R of party R. (Note that the citizens do not observe message m_A of party A. The reason for this will become clear in the next subsection.) Given prior F, message m_R, and private type t, each citizen derives the posterior expected state x, and chooses which party to support.

B. Sequential Disclosure and Sequential Obfuscation

We consider two variants of the basic setting: a model of sequential disclosure and a model of sequential obfuscation. These models impose different constraints on the strategy of party R.

To define sequential disclosure and sequential obfuscation, we introduce the following notation. Because the parties' utilities depend only on the citizens' total support, which in turn depends only on the expected state, the information disclosed by a message m can be summarized by the posterior expected state induced by this message. Given a pair of strategies (ϕ_A, ϕ_R), let $\mu_A(\phi_A) \in \Delta(\Theta)$ be the distribution of the expected state induced by observation of messages of party A, and let $\mu_R(\phi_A, \phi_R) \in \Delta(\Theta)$ be the distribution of the expected state induced by observation of messages of party R.

We compare distributions of the expected state by their Blackwell informativeness (David Blackwell, 1953) for the citizens. We say that distribution μ' is *more informative* than distribution μ'', denoted by $\mu' \succeq \mu''$, if μ' is a mean preserving spread of μ''.

Sequential disclosure. In sequential disclosure, party R reveals information in addition to what has been revealed by party A's message m_A. That is, the citizens can always deduce m_A from m_R. This formalism captures the idea that the citizens observe both messages, and party R cannot hide what has been revealed by party A. By Blackwell (1953), this means that, given the distribution $\mu_A(\phi_A)$ of the expected state induced by party A's strategy ϕ_A, strategy ϕ_R must induce a weakly more informative distribution, but no more informative than fully revealing the state, so ϕ_R must satisfy

$$F \succeq \mu_R(\phi_A, \phi_R) \succeq \mu_A(\phi_A).$$

Sequential obfuscation. In sequential obfuscation, party R obfuscates (or garbles) information revealed by party A's message. That is, if the citizens were able to observe m_A instead of m_R, they could deduce m_R. This means that, given the distribution $\mu_A(\phi_A)$ of the expected state induced by party A's strategy ϕ_A, strategy ϕ_R must induce a weakly less informative distribution, so ϕ_R must satisfy

$$\mu_A(\phi_A) \succeq \mu_R(\phi_A, \phi_R).$$

We are interested in the characterisation and comparison of equilibria in the models of sequential disclosure and sequential obfuscation. The solution concept is subgame perfect equilibrium.

III. Equilibrium Outcomes

Given a posterior expected value x of the state, the citizen with type $t = x$ is indifferent between supporting the accuser and the respondent. The fractions of the population that support the accuser and the responder are the masses of types below x and above x, respectively, so they are equal to $G(x)$ and $1 - G(x)$. Define

$$V_A(x) = u_A(G(x)) \quad \text{and} \quad V_R(x) = u_R(1 - G(x)). \tag{1}$$

So, $V_i(x)$ is party i's utility when the posterior expected value of the state is x, $i = A, R$. Note that $V_A(x)$ is increasing and $V_R(x)$ is decreasing in x. We will refer to $V_i(x)$ as party i's *indirect utility*.

An *outcome* μ of sequential disclosure or sequential obfuscation with a given pair of strategies (ϕ_A, ϕ_R) is the distribution of the posterior expected state induced by the message of party R, $\mu = \mu_R(\phi_A, \phi_R)$. The outcome summarizes the information revealed to the citizens. It also determines the expected utilities of the parties. Let $V_i(\mu)$ be the expected utility of party i when the outcome is $\mu \in \Delta(\Theta)$,

$$V_i(\mu) = \int_{x \in \Theta} V_i(x) \mathrm{d}\mu(x), \quad i = A, R.$$

Note that outcomes and the associated expected utilities are not affected by zero probability events. That is, two pairs of strategies (ϕ_A, ϕ_R) and (ϕ'_A, ϕ'_R) that send the same messages

with probability one lead to the same outcome.

Given a prior distribution F, an outcome $\mu \in \Delta(\Theta)$ is *feasible* if it can be obtained by an information structure, that is, if F is more informative than μ (Blackwell, 1953). Let $\mathcal{M}$ be the set of feasible outcomes,

$$\mathcal{M} = \{\mu \in \Delta(\Theta) : F \succeq \mu\}.$$

We use the notion of unimprovable outcomes[3] to simplify the problems of finding subgame perfect equilibria in sequential disclosure and sequential obfuscation.

A feasible outcome $\mu \in \mathcal{M}$ is *unimprovable by disclosure* for party R if that party cannot be better off with any outcome μ' that can be obtained from μ by disclosure,

$$V_R(\mu) \geq V_R(\mu') \text{ for all } \mu' \in \mathcal{M} \text{ such that } \mu' \succeq \mu.$$

An outcome $\mu \in \mathcal{M}$ is *unimprovable by obfuscation* for party R if that party cannot be better off with any outcome μ' that can be obtained from μ by obfuscation,

$$V_R(\mu) \geq V_R(\mu') \text{ for all } \mu' \in \mathcal{M} \text{ such that } \mu \succeq \mu'.$$

Let $\mathcal{M}_R^D$ and $\mathcal{M}_R^O$ be the set of feasible outcomes that are unimprovable by disclosure and obfuscation, respectively, for party R.

We now show that the problem of sequential disclosure (sequential obfuscation) is equivalent to the problem where only party A chooses an information structure. Because party R is able to distort some choices of party A by revealing (obfuscating) information, party A can only attain outcomes that party R does not want to improve upon. Party A then chooses the best among such outcomes.

Consider two problems where party A chooses an outcome to maximize her expected payoff among the outcomes that are unimprovable by disclosure and obfuscation, respectively, for party R:

$$(\mathrm{P_D}) \qquad \max_{\mu \in \mathcal{M}_R^D} V_A(\mu),$$

$$(\mathrm{P_O}) \qquad \max_{\mu \in \mathcal{M}_R^O} V_A(\mu).$$

Observation 1. *An outcome $\mu \in \Delta(\Theta)$ is an equilibrium outcome of sequential disclosure (sequential obfuscation) if and only if it is a solution of problem* $(\mathrm{P_D})$ *(respectively,* $(\mathrm{P_O})$*).*

Li and Norman (2021) prove the statement of Observation 1 for sequential disclosure, and Lipnowski, Mathevet and Wei (2020) prove it for sequential obfuscation.[4] The idea behind

[3]Variants of this notion appear in Gentzkow and Kamenica (2017*b*) and Li and Norman (2021).

[4]Note that the setting in Lipnowski, Mathevet and Wei (2020) is set in a very different context than the setting in our paper, but it can be interpreted as sequential obfuscation. In their paper, a single sender communicates with a single receiver, where the latter has costly attention and is willing to obtain coarser information when fine details

Observation 1 is reminiscent of the revelation principle. If an equilibrium of sequential obfuscation by two parties leads to an outcome μ, then it must remain equilibrium if party A implements μ directly. Party R then has no incentive to obfuscate μ, because if it did, it would have done so in the original equilibrium.

Observation 2. *For generic utility functions u_A and u_R, the games of sequential disclosure and sequential obfuscation have unique equilibrium outcomes.*

Li and Norman (2021) prove the statement of Observation 2 for sequential disclosure, and the proof for sequential obfuscation is analogous. To see why Observation 2 holds in the case of sequential obfuscation, by Observation 1 we only need to consider the problem ($\mathrm{P_O}$). Party A maximizes the linear functional $\int_{x\in\Theta} u_A(G(x))\mathrm{d}\mu(x)$ over the set $\mathcal{M}_R^O$. Because $\mathcal{M}_R^O$ is independent of u_A, if there is more than one maximizer for party A, the ties can be broken by a slight perturbation of u_A.

The above observations illuminate the difference between disclosure and obfuscation. Loosely speaking, sequential disclosure restricts party A's choice to outcomes that are sufficiently revealing from party R's perspective, so that party R does not wish to reveal any more. Similarly, sequential obfuscation restricts party A's choice to outcomes that are sufficiently unrevealing from party R's perspective, so that party R does not wish to obfuscate them. The set of outcomes that are unimprovable by both disclosure and obfuscation for party R has measure zero set for a generic distribution of citizens G. Thus, party A optimizes on two essentially disjoint sets in the two problems, one clearly favoring more information disclosure than the other.

Let us now support the above argument by a formal result. It demonstrates that sequential obfuscation cannot be more informative than sequential disclosure.

PROPOSITION 1: *Let μ^D and μ^O be equilibrium outcomes of sequential disclosure and sequential obfuscation, respectively, and suppose that the parties' expected utilities are not identical,*

$$\big(V_A(\mu^D), V_R(\mu^D)\big) \neq \big(V_A(\mu^O), V_R(\mu^O)\big)\,.$$

Then μ^O cannot be more informative than μ^D.

The proof is in Appendix B.

Next, we establish the conditions for when the equilibrium outcome fully reveals the state, and when it reveals nothing.

An outcome μ is called *no disclosure* if it reveals no information about the state, that is, μ puts probability one on the prior expected value of the state.

An outcome μ is called *full disclosure* if it reveals the state, that is, $\mu = F$.

When comparing two outcomes μ' and μ'', we say that μ' *Pareto dominates* μ'' if

$$\big(V_A(\mu'), V_R(\mu')\big) \gneqq \big(V_A(\mu''), V_R(\mu'')\big)\,.$$

An outcome μ is *Pareto undominated* if there is no other feasible outcome that Pareto dominates μ. An outcome μ is *Pareto dominant* if it is Pareto dominates all other feasible

are not worth the cost, that is, to obfuscate information.

outcomes.

We show that sequential disclosure typically reveals some information, except when both parties unanimously prefer to reveal nothing. Moreover, it fully reveals the state if at least one party prefers to do so.

PROPOSITION 2: *In the sequential disclosure game:*

(i) If full disclosure is Pareto undominated, then it is an equilibrium outcome.

(ii) If no disclosure is Pareto dominant, then it is an equilibrium outcome. If no disclosure is not Pareto dominant, then, for generic preferences, it is not an equilibrium outcome.

The proof is in Appendix C.

Intuitively, part (i) follows from the observation that in the sequential disclosure game, full disclosure is unilaterally enforceable by each party. Thus, if every outcome is worse than full disclosure for at least one party, then full disclosure becomes is the only outcome that can emerge in equilibrium. Part (ii) holds by Observation 1 and the fact that when no disclosure Pareto dominates all other feasible outcomes, then it must be unimprovable for party R and most preferred for party A. Conversely, when no disclosure does not Pareto dominate all other feasible outcomes, then either it is improvable for party R, or, for generic preferences, it is strictly inferior to some outcome for party A.

Symmetrically, sequential obfuscation typically obfuscates some information, except when both parties unanimously prefer to fully reveal the state. Moreover, it reveals nothing if at least one party prefers to do so.

PROPOSITION 3: *In the sequential obfuscation game:*

(i) If no disclosure is Pareto undominated, then it is an equilibrium outcome.

(ii) If full disclosure is Pareto dominant, then it is an equilibrium outcome. If full disclosure is not Pareto dominant, then, for generic preferences, it is not an equilibrium outcome.

The intuition and proof are symmetric to those for Proposition 2.

IV. Special Cases

In this section we examine some notable special cases. For these cases, we show that either sequential disclosure fully reveals the state, or sequential obfuscation reveals no information, or both are true at the same time. Thus, sequential disclosure is more Blackwell informative than sequential obfuscation in these cases. Importantly, and perhaps surprisingly, these results are completely independent of the prior distribution of the state, so they hold regardless how much the prior favors the accuser or the respondent.

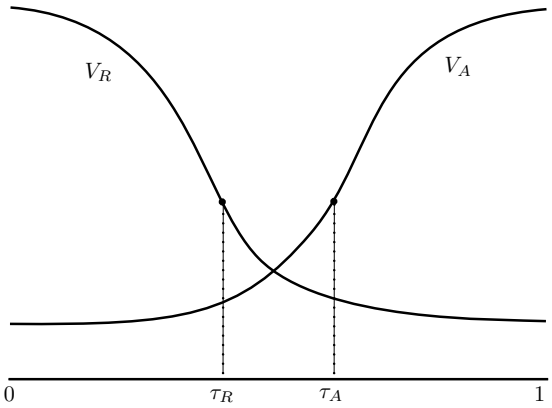

Figure 1. : S-shaped V_A and inverted S-shaped V_R.

A. Log-concave preferences

For several results of this section, we will assume that the distribution of citizens' types G has a strictly log-concave density g. Formally,

(A_1) G admits a continuously differentiable density g, and $\ln g(\cdot)$ is strictly concave on $[0,1]$.

Log-concavity of a probability density is a common assumption in a variety of economic applications, such as voting, signalling, and monopoly pricing (see Section 7 in Bagnoli and Bergstrom, 2005). Log-concave densities exhibit nice properties, such as unimodality and hazard rate monotonicity. Many familiar probability density functions are log-concave (see Table 1 in Bagnoli and Bergstrom, 2005).

In addition, for several results of this section, we assume that the marginal utilities of both parties, u'_A and u'_R, are log-concave, so

(A_2) $\ln(u'_A(\cdot))$ and $\ln(u'_R(\cdot))$ are concave on $[0,1]$.

A log-concave marginal utility of party $i = A, R$ represents the preferences whose Arrow-Pratt coefficient of risk aversion $-u''_i(y)/u'_i(y)$ is increasing. In words, the more support a party gains, the less it likes to gamble with this support. Also, a log-concave marginal utility function is monotone or single-peaked. Thus, this assumption includes the case relevant in political applications in which the parties care more about obtaining the support of the citizens near the median of the population distribution and less about those at the extremes. This is the case when the parties can be interested in reaching the support by the simple majority, so each $u_i(y)$ smoothly approximates the step function with value 0 when $y < 1/2$ and 1 when $y > 1/2$.

Under assumptions (A_1) and (A_2), the indirect utilities V_A and V_R have specific shapes,

which play a role in some of the results presented below. As illustrated in Fig. 1, $V_A(x)$ is *strictly S-shaped*, that is, it is strictly convex up to an inflexion point, denoted by τ_A, and then strictly concave. Symmetrically, $V_R(x)$ is *strictly inverted S-shaped*, that is, it is strictly concave up to an inflexion point, denoted by τ_R, and then strictly convex.

LEMMA 1: *Suppose that assumptions* (A_1) *and* (A_2) *hold. There exist* $\tau_A, \tau_R \in [0,1]$ *such that*

(i) $V_A(x)$ *is strictly convex for* $x < \tau_A$ *and strictly concave for* $x > \tau_A$*;*

(ii) $V_R(x)$ *is strictly concave for* $x < \tau_R$ *and strictly convex for* $x > \tau_R$*.*

The proof is in Appendix D.

B. Constant marginal utility ratio. Risk neutrality and zero-sum preferences

Let q be the mass of citizens who support party A, so $1-q$ is the mass of citizens who support party R. We say that utility functions u_A and u_R have *constant marginal utility ratio (CMUR)* if

$$\frac{u'_A(q)}{u'_R(1-q)} \text{ is constant.}$$

This condition includes two special cases that are prominent in the literature:

(i) when the preferences are zero-sum or constant-sum;

(ii) when the preferences are linear, so the parties are risk neutral.

Observe that constant marginal utility ratio can be equivalently expressed as

$$u_R(1-q) = b - cu_A(q) \text{ for some } b \in \mathbb{R} \text{ and } c > 0, \tag{2}$$

so the utilities are linear functions of each other. An immediate consequence of (2) is that, by (1), the expected indirect utilities from any outcome μ satisfy $V_R(\mu) = b - cV_A(\mu)$. Thus, for any two outcomes μ' and μ'' we have

$$V_A(\mu') \geq (>) V_A(\mu'') \iff V_R(\mu') \leq (<) V_R(\mu''). \tag{3}$$

In words, CMUR generalizes the idea of zero-sum preferences, because it implies that there is no room for cooperation: what is better for one is always worse for the other.

When the utility functions satisfy CMUR, the difference between equilibrium outcomes of sequential disclosure and sequential obfuscation is extreme: the former fully reveals the state and the latter reveals no information at all.

THEOREM 1: *Suppose that* $u'_A(q)/u'_R(1-q)$ *is constant. Then full disclosure (no disclosure) is an equilibrium outcome of sequential disclosure (sequential obfuscation, respectively). Moreover, these equilibrium outcomes are unique in the respective games if assumptions* (A_1) *and* (A_2) *are satisfied.*

The proof is in Appendix E.

To see why Theorem 1 holds, notice that condition (3) implies that both full disclosure and no disclosure are Pareto undominated. Thus, by Propositions 2(i) and 3(i), full disclosure is an equilibrium outcome of the sequential disclosure game, and no disclosure is an equilibrium outcome of the sequential obfuscation game. However, we cannot claim the uniqueness of these equilibria, not even generically, because under CMUR, u_A cannot be perturbed independently of u_R, so Observation 2 does not apply. Yet, the structure imposed by assumptions (A_1) and (A_2) is sufficient to ensure uniqueness.

REMARK 1: *In the zero-sum-like situation stipulated by CMUR, one could expect that the second mover has an advantage. However, as apparent from Theorem 1, when CMUR holds, the order of moves plays no role in the sequential obfuscation and sequential disclosure games.*

C. Decreasing marginal utility ratio. Risk aversion.

We say that utility functions u_A and u_R have *decreasing marginal utility ratio (DMUR)* if

$$\frac{u'_A(q)}{u'_R(1-q)} \quad \text{is decreasing.}$$

DMUR means that every utility unit gained by one party translates into an increasingly larger number of utility units lost by the other party. That is, DMUR can be expressed as[5]

$$u_A(1-u_R^{-1}(y)) \text{ is concave in } y. \tag{4}$$

This condition incorporates the case of risk averse preferences, that is, it holds when both u_A and u_R are concave.

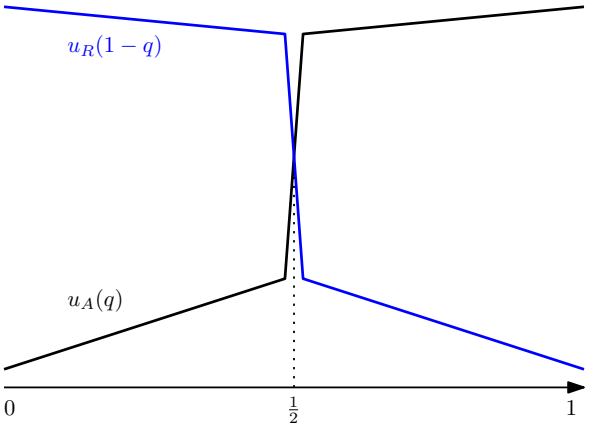

Figure 2. : Decreasing marginal utility ratio $u'_A(q)/u'_R(1-q)$.

DMUR can be seen as a feature of democratic regimes where the minority party stands to gain more from increasing its support than the majority party stands to lose. For

[5]Let the utility of R be y, so $y = u_R(1-q)$. Then $q = 1-u_R^{-1}(y)$, and $u_A(q) = u_A(1-u_R^{-1}(y))$.

illustration, consider the utility functions depicted in Fig. 2. To highlight the property of decreasing ratio of marginal utilities, the illustrated functions are piecewise linear. Initially, as the support q of party A increases, party A's utility (black line) grows faster than party R's utility (blue line) declines. Around the simple majority threshold, $q = 1/2$, there is a sharp increase in party A's utility and a quantitatively identical decrease in party R's utility. Then, as the support of party A continues to grow after it has secured the majority, it gains less utility from each additional supporter than party R loses. In simple words, gaining additional support is more valuable when you do not have the majority than when you do.

We now show that under the condition of decreasing marginal utility ratio, sequential obfuscation leads to no information disclosed about the state.

THEOREM 2: *Suppose that $u'_A(q)/u'_R(1-q)$ is decreasing. Then no disclosure is an equilibrium outcome of sequential obfuscation. Moreover, it is the unique equilibrium outcome if either $u'_A(q)/u'_R(1-q)$ is strictly decreasing, or assumptions* (A_1) *and* (A_2) *are satisfied.*

The proof is in Appendix F.

The intuition for the no disclosure result under sequential obfuscation is as follows. Condition (4) means that one party's utility is concave when the unit of measurement is the other party's utility unit. So, each party has a diminishing incentive to fight for the support of another citizen. This leads to the situation where for every informative disclosure mechanism, at least one or the parties would benefit from garbling information, and no disclosure is Pareto undominated. Thus, by Proposition 3(i), no disclosure is an equilibrium outcome. The additional assumptions, either strict DMUR or (A_1)–(A_2), add strict curvature of the indirect utility functions V_A and V_R, which is sufficient to guarantee the uniqueness of the equilibrium outcome.

Let us make a conclusion from Theorem 2 in the context of our political story. In the case of democratic regimes where the minority party stands to gain more from increasing its support than the majority party stands to lose, toxic debates (sequential obfuscation) reveal nothing about the state, whereas constructive debates (sequential disclosure) reveal some information, except when no disclosure is Pareto dominant. This provides a possible explanation why democratic regimes are often not very good in digging out the truth, and highlights the danger of negative criticism and contempt for truth discovery in political debates.

Note that if the utility functions u_A and u_R are concave, so that both parties prefer the prior expected support with certainty to any distribution over the citizens' support, it does not automatically imply that any type of communication game leads to no disclosure. In fact, by Proposition 2(ii), the sequential disclosure game leads to revelation of some information (except when no disclosure is Pareto dominant), and it can even lead to full disclosure. For example, let

$$u_A(y) = u_R(y) = \sqrt{y} \quad \text{and} \quad G(x) = 1 - e^{-x}.$$

Then $V_A(x) = u_A(G(x)) = \sqrt{1-e^{-x}}$ is strictly concave in x, whereas $V_R(x) = u_R(1 - G(x)) = \sqrt{e^{-x}}$ is strictly convex in x. This means that the unique most preferred outcome of party R is full disclosure (e.g., Kamenica and Gentzkow, 2011). Thus, by Proposition 2(i), in the sequential disclosure game, the unique equilibrium outcome is full disclosure,

despite both parties being strictly risk averse.

D. *Increasing marginal utility ratio. Risk seeking.*

We say that utility functions u_A and u_R have *increasing marginal utility ratio (IMUR)* if

$$\frac{u_A'(q)}{u_R'(1-q)} \text{ is increasing.}$$

Symmetrically to DMUR, IMUR means that every utility unit lost by one party translates into an increasingly larger number of utility units gained by the other party. That is, IMUR can be expressed as

$$u_A(1-u_R^{-1}(y)) \text{ is convex in } y. \tag{5}$$

This condition incorporates the case of risk seeking preferences, that is, it holds when both u_A and u_R are convex.

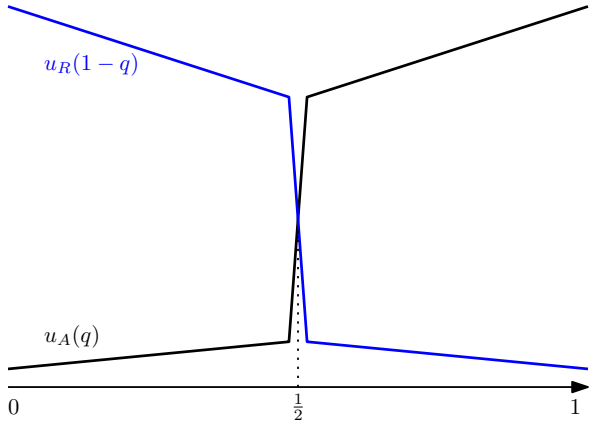

Figure 3. : Increasing marginal utility ratio $u_A'(q)/u_R'(1-q)$.

IMUR can be seen as a feature of authoritarian regimes where the majority party stands to gain more from squashing the minority opposition than the opposition stands to lose. In Fig. 3, which shows u_A and u_R that satisfy IMUR, the situation is the opposite to DMUR: party A gains relatively more when it is supported by the majority than when it is supported by the minority. In simple words, gaining additional support is more valuable when you have the majority than when you do not.

We now show that under the condition of increasing marginal utility ratio and log-concave preferences, sequential disclosure fully reveals the state.

THEOREM 3: *Suppose that $u_A'(q)/u_R'(1-q)$ is increasing and assumptions* (A_1) *and* (A_2) *are satisfied. Then full disclosure is the unique equilibrium outcome of sequential disclosure.*

The proof is in Appendix G.

The intuition for Theorem 3 is different from that for Theorem 2. Unlike in the case of DMUR, IMUR cannot guarantee that full disclosure is Pareto undominated (see the counterexample in Section IV.E), and thus Proposition 2(i) does not apply. To prove Theorem 3, we rely on the additional structure due to the log-concavity assumptions (A_1) and (A_2) as follows.

Recall that, given (A_1) and (A_2), function V_A is S-shaped, and function V_R is inverted S-shaped. We establish that under IMUR, there is an overlap of the intervals on which V_A and V_R are convex, that is, the inflection points satisfy $\tau_R \leq \tau_A$, as illustrated in Fig. 1. Now consider any message that does not reveal the state exactly and induces a posterior expected value of the state x. In the neighborhood of x, either $V_A(x)$ or $V_R(x)$ or both are locally convex. So at least one party will benefit from applying a mean preserving spread to the posterior x in a small enough neightborhood and be strictly better off, because of the convexity of the utility function. So, messages in equilibrium must be revealing about the state.

Let us make a conclusion from Theorem 3 in the context of our political story. In the case of authoritarian regimes where the majority party stands to gain more from squashing the minority opposition than the opposition stands to lose, constructive debates (sequential disclosure) reveal the state, whereas toxic debates (sequential obfuscation) do not, except when full disclosure is Pareto dominant. Perhaps, this is why authoritarian regimes, which understand the threat of constructive debates in exposing the truth, are so keen to discredit or completely shut down the opposition.

E. Counterexample

No disclosure is Pareto undominated under DMUR (as shown in the proof of Theorem 2). But the symmetric claim, that full disclosure is Pareto undominated under IMUR, need not be true. It is only true if the prior F has support on two values of the state.

For a counterexample, let F be uniform on $[0,1]$, and let

$$u_A(y) = u_R(y) = y^2 \quad \text{and} \quad G(x) = \begin{cases} 0 & \text{if } x \in [0,1/3], \\ 1/2 & \text{if } x \in (1/3,2/3], \\ 1 & \text{if } x \in (2/3,1]. \end{cases}$$

Then $V_A(x) = u_A(G(x))$ and $V_R(x) = u_R(1-G(x))$ are as shown in Fig. 4. Let us compare the full disclosure and the cutoff disclosure $\mu_{1/2}$ that reveals whether the state is above or below 1/2. Observe that $\mu_{1/2}$ induces the posteriors 1/4 and 3/4 equally likely, and yields the expected utility of 1/2 for both parties (illustrated by the midpoint of dashed lines in Fig. 4). However, full disclosure yields for each $i = A, R$ the expected utility

$$\int_0^1 V_i(x)\mathrm{d}x = \frac{1}{3}\cdot 0 + \frac{1}{3}\cdot\frac{1}{4} + \frac{1}{3}\cdot 1 = \frac{5}{12} < \frac{1}{2}.$$

That is, both parties strictly prefer $\mu_{1/2}$ to full disclosure.

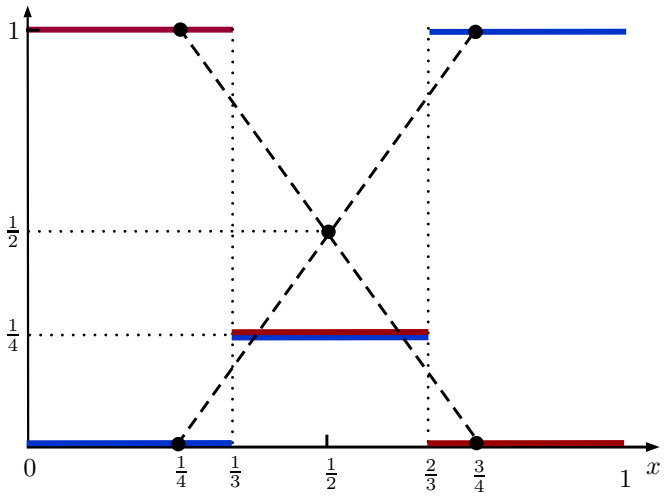

Figure 4. : Indirect utilities $V_A(x)$ of party A (solid blue) and $V_R(x)$ of party R (solid red).

V. Discussion

Let us discuss two assumptions in our model that deserve particular attention: the commitment to information strategies before learning the state, and the sequentiality of moves of the parties.

Ex-ante symmetric information and commitment. As standard in Bayesian persuasion literature, we assume that when making their choices, the parties and the audience are symmetrically informed about the state and fully committed to their strategies. That is, the parties choose their information disclosure (or obfuscation) strategies before learning anything about the state that the audience does not already know, and then they send messages according to the chosen strategies.

One aspect of the ex-ante commitment assumption is that communication is not cheap talk. That is, the parties can hide or garble information, but they cannot alter their messages on the go and they cannot outright lie to the citizens. This is justifiable when the parties have their reputation to maintain, and there are substantial penalties to one's reputation for lying. We also know from the literature that Bayesian persuasion results are robust to minor departures from full commitment (Elliot Lipnowski, Doron Ravid and Denis Shishkin, 2022; Yingni Guo and Eran Shmaya, 2021; Daehong Min, 2021; Ran Eilat and Zvika Neeman, 2023).

The other aspect of the ex-ante commitment assumption is that the parties do not have private information about the state when choosing their information disclosure strategies. While this is a foundational assumption in the Bayesian persuasion literature, one could argue that in practice the parties can be privately informed about the state and play a signaling game where they signal about the state by their choices of information structures. For a justification of our assumption, consider the following arguments.

First, this paper aspires to contribute to a broader literature on information design and Bayesian persuasion, by contrasting disclosure and garbling in sequential Bayesian persuasion.

Second, ex-ante commitment by the parties to information structures leads to the same outcome as a pooling equilibrium of the signaling game, where the parties choose the very same information structures regardless of their private information. Pooling equilibria are sustainable as sequential equilibria of the signaling game under an additional mild assumption that the parties' information structures cannot be perfectly accurate (but they can be arbitrarily close to perfectly accurate). Specifically, a pooling equilibrium can be supported by assuming that if any party deviates from the prescribed path, the audience will adopt the extreme posterior that maximally favors the deviant's opponent party, thus negating any potential benefit for the deviant. This construction is used by Andriy Zapechelnyuk (2023) to demonstrate the equivalence of implementable outcomes by the uninformed and informed sender in a standard single sender-receiver model, and it straightforwardly extends to the setting of this paper.

Lastly, our paper focuses on characterizing the cases where the extreme equilibrium outcomes – full disclosure and no disclosure – emerge. We conjecture that these results are robust to the introduction of private information of the parties about the state. The key feature of these results is that a particular outcome (full disclosure or no disclosure) is unilaterally enforceable and, independently of the distribution of the state, at least one party wishes to enforce it. Informative signaling by the first mover benefits the second mover, but, because the parties have conflicting interests, it can only hurt the first mover. Consequently, analogously to the no-trade theorem of Paul Milgrom and Nancy Stokey (1982), the first mover cannot gain by signaling. The equilibrium of the signaling game is pooling, and its outcome is the only unilaterally enforceable outcome.

Sequential moves. We assume that the parties choose their information structures sequentially. On the one hand, it is plausible that the respondent chooses her strategy only after she has seen the choice of the accuser. On the other hand, the sequence of moves does not change the outcome if it is full disclosure or no disclosure, which are focal cases for this paper.

From the technical perspective, sequential moves allow us to substantially reduce the set of equilibria, and to obtain a unique equilibrium outcome for generic preferences. It is known that simultaneous disclosure of an initially hidden state leads to a plethora of equilibria, because any outcome that either party cannot improve by additional disclosure is an equilibrium (Gentzkow and Kamenica, 2017*b*). For instance, full disclosure is always an equilibrium. Symmetrically, simultaneous obfuscation of an initially revealed state leads to a plethora of equilibria, e.g., no disclosure is always an equilibrium. One needs to resort to equilibrium refinements, such as a strictly positive cost of information disclosure, to obtain a meaningful result. However, when the parties move sequentially, the first mover has the power to select among multiple outcomes that neither player can unilaterally improve upon. Thus, with sequential moves, the multiplicity of equilibria is much less of an issue, and generically a unique equilibrium outcome is obtained.

References

Alonso, Ricardo, and Odilon Câmara. 2016. "Political disagreement and information in elections." *Games and Economic Behavior*, 100: 390–412.

Ambrus, Attila, and Satoru Takahashi. 2008. "Multi-sender cheap talk with restricted state spaces." *Theoretical Economics*, 3: 1–27.

Ambrus, Attila, and Shih En Lu. 2014. "Almost fully revealing cheap talk with imperfectly informed senders." *Games and Economic Behavior*, 88: 174–189.

Arieli, Itai, Yakov Babichenko, and Fedor Sandomirskiy. 2022. "Bayesian persuasion with mediators." Available at arXiv:2203.04285.

Arieli, Itai, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita. 2023. "Optimal persuasion via bi-pooling." *Theoretical Economics*, 18: 15–36.

Au, Pak Hung, and Keiichi Kawai. 2020. "Competitive information disclosure by multiple senders." *Games and Economic Behavior*, 119: 56–78.

Austen-Smith, David. 1990. "Information transmission in debate." *American Journal of Political Science*, 34: 124–152.

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

Bagnoli, Mark, and Ted Bergstrom. 2005. "Log-concave probability and its applications." *Economic Theory*, 26: 445–469.

Battaglini, Marco. 2002. "Multiple referrals and multidimensional cheap talk." *Econometrica*, 70: 1379–1401.

Battaglini, Marco. 2004. "Policy advice with imperfectly informed experts." *The BE Journal of Theoretical Economics*, 4: Article 1.

Blackwell, David. 1953. "Equivalent comparisons of experiments." *Annals of Mathematical Statistics*, 24: 265–272.

Bloedel, Alexander W, and Ilya Segal. 2020. "Persuading a rationally inattentive agent." Available at SSRN 3164033.

Boleslavsky, Raphael, and Christopher Cotton. 2018. "Limited capacity in project selection: Competition through evidence production." *Economic Theory*, 65: 385–421.

Brocas, Isabelle, Juan D Carrillo, and Thomas R Palfrey. 2012. "Information gatekeepers: Theory and experimental evidence." *Economic Theory*, 51: 649–676.

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang. 2019. "Pivotal persuasion." *Journal of Economic theory*, 180: 178–202.

D'Agostino, Elena, and Daniel J Seidmann. 2022. "The order of presentation in trials: Plaintive plaintiffs." *Games and Economic Behavior*, 132: 328–336.

Duffie, Darrell, Piotr Dworczak, and Haoxiang Zhu. 2017. "Benchmarks in search markets." *Journal of Finance*, 72: 1983–2044.

Dworczak, Piotr, and Giorgio Martini. 2019. "The simple economics of optimal persuasion." *Journal of Political Economy*, 127: 993–2048.

Edmond, Chris, and Yang K Lu. 2021. "Creating confusion." *Journal of Economic Theory*, 191: 105145.

Eilat, Ran, and Zvika Neeman. 2023. "Communication with endogenous deception costs." *Journal of Economic Theory*, 207: 105572.

Gehlbach, Scott, and Konstantin Sonin. 2014. "Government control of the media." *Journal of Public Economics*, 118: 163–171.

Gentzkow, Matthew, and Emir Kamenica. 2016. "A Rothschild-Stiglitz approach to Bayesian persuasion." *American Economic Review: Papers & Proceedings*, 106: 597–601.

Gentzkow, Matthew, and Emir Kamenica. 2017*a*. "Bayesian persuasion with multiple senders and rich signal spaces." *Games and Economic Behavior*, 104: 411–429.

Gentzkow, Matthew, and Emir Kamenica. 2017*b*. "Competition in persuasion." *Review of Economic Studies*, 85: 300–322.

Gilligan, Thomas W, and Keith Krehbiel. 1989. "Asymmetric information and legislative rules with a heterogeneous committee." *American Journal of Political Science*, 33: 459–490.

Ginzburg, Boris. 2019. "Optimal information censorship." *Journal of Economic Behavior and Organization*, 163: 377–385.

Gitmez, Arda, and Pooya Molavi. 2022. "Polarization and media bias." Available at arXiv:2203.12698.

Glazer, Jacob, and Ariel Rubinstein. 2001. "Debates and decisions: On a rationale of argumentation rules." *Games and Economic Behavior*, 36: 158–173.

Goldstein, Itay, and Yaron Leitner. 2018. "Stress tests and information disclosure." *Journal of Economic Theory*, 177: 34–69.

Gul, Faruk, and Wolfgang Pesendorfer. 2012. "The war of information." *Review of Economic Studies*, 79: 707–734.

Guo, Yingni, and Eran Shmaya. 2021. "Costly miscalibration." *Theoretical Economics*, 16: 477–506.

Hu, Ju, and Zhen Zhou. 2022. "Disclosure in epidemics." *Journal of Economic Theory*, 202: 105469.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review*, 101: 2590–2615.

Kleiner, Andreas, Benny Moldovanu, and Philipp Strack. 2021. "Extreme points and majorization: Economic applications." *Econometrica*, 89: 1557–1593.

Koessler, Frederic, Marie Laclau, Jérôme Renault, and Tristan Tomala. 2022. "Long information design." *Theoretical Economics*, 17: 883–927.

Kolotilin, Anton. 2015. "Experimental design to persuade." *Games and Economic Behavior*, 90: 215–226.

Kolotilin, Anton. 2018. "Optimal information disclosure: A linear programming approach." *Theoretical Economics*, 13: 607–636.

Kolotilin, Anton, and Andriy Zapechelnyuk. 2019. "Persuasion Meets Delegation." mimeo.

Kolotilin, Anton, Tymofiy Mylovanov, and Andriy Zapechelnyuk. 2022. "Censorship as optimal persuasion." *Theoretical Economics*, 17: 561–585.

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. 2017. "Persuasion of a privately informed receiver." *Econometrica*, 85: 1949–1964.

Krishna, Vijay, and John Morgan. 2001*a*. "Asymmetric information and legislative rules: Some amendments." *American Political Science Review*, 95: 435–452.

Krishna, Vijay, and John Morgan. 2001*b*. "A model of expertise." *Quarterly Journal of Economics*, 116: 747–775.

Levy, Gilat, and Ronny Razin. 2007. "On the limits of communication in multidimensional cheap talk: A comment." *Econometrica*, 75: 885–893.

Levy, Gilat, and Ronny Razin. 2012. "When do simple policies win?" *Economic Theory*, 49: 621–637.

Li, Fei, and Peter Norman. 2018. "On Bayesian persuasion with multiple senders." *Economics Letters*, 170: 66–70.

Li, Fei, and Peter Norman. 2021. "Sequential persuasion." *Theoretical Economics*, 16: 639–675.

Li, Fei, Yangbo Song, and Mofei Zhao. 2023. "Global manipulation by local obfuscation." *Journal of Economic Theory*, 207: 105575.

Li, Ming. 2010. "Advice from multiple experts: A comparison of simultaneous, sequential, and hierarchical communication." *The BE Journal of Theoretical Economics*, 10: Article 18.

Lipman, Barton L, and Duane J Seppi. 1995. "Robust inference in communication games with partial provability." *Journal of Economic Theory*, 66: 370–405.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin. 2022. "Persuasion via weak institutions." *Journal of Political Economy*, 130: 2705–2730.

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei. 2020. "Attention management." *American Economic Review: Insights*, 2: 17–32.

Milgrom, Paul, and Nancy Stokey. 1982. "Information, trade and common knowledge." *Journal of economic theory*, 26: 17–27.

Min, Daehong. 2021. "Bayesian persuasion under partial commitment." *Economic Theory*, 72: 743–764.

Mylovanov, Tymofiy, and Andriy Zapechelnyuk. 2013*a*. "Decision rules revealing commonly known events." *Economics Letters*, 119: 8–10.

Mylovanov, Tymofiy, and Andriy Zapechelnyuk. 2013*b*. "Optimal arbitration." *International Economic Review*, 54: 769–785.

Orlov, Dmitry, Pavel Zryumov, and Andrzej Skrzypach. 2022. "Design of macroprudential stress tests." Available at SSRN 2977016.

Ottaviani, Marco, and Peter Sørensen. 2001. "Information aggregation in debate: Who should speak first?" *Journal of Public Economics*, 81: 393–421.

Ravindran, Dilip, and Zhihan Cui. 2022. "Competing persuaders in zero-sum games." Available at SSRN 4241719.

Romanyuk, Gleb, and Alex Smolin. 2019. "Cream skimming and information design in matching markets." *American Economic Journal: Microeconomics*, 11: 250–276.

Shin, Hyun Song. 1998. "Adversarial and inquisitorial procedures in arbitration." *RAND Journal of Economics*, 378–405.

Spector, David. 2000. "Rational debate and one-dimensional conflict." *Quarterly Journal of Economics*, 115: 181–200.

Spiegler, Ran. 2006. "Argumentation in multi-issue debates." *Social Choice and Welfare*, 26: 385–402.

Vatter, Benjamin. 2022. "Quality disclosure and regulation: Scoring design in medicare advantage." Available at SSRN 4250361.

Wolinsky, Asher. 2002. "Eliciting information from multiple experts." *Games and Economic Behavior*, 41: 141–160.

Wu, Wenhao. 2022. "Sequential bayesian persuasion." Available at SSRN 3841869.

Zapechelnyuk, Andriy. 2013. "Eliciting information from a committee." *Journal of Economic Theory*, 148: 2049–2067.

Zapechelnyuk, Andriy. 2020. "Optimal quality certification." *American Economic Review: Insights*, 2: 161–176.

Zapechelnyuk, Andriy. 2023. "On the equivalence of information design by uninformed and informed principals." *Economic Theory*, forthcoming.

Appendix

Proof of Proposition 1

By contradiction, suppose that $\mu^O \succeq \mu^D$. Then μ^O is attainable by disclosure from μ^D, and μ^D is attainable by obfuscation from μ^O. Thus, $\mu^D \in \mathcal{M}_R^O$ and $\mu^O \in \mathcal{M}_R^D$. By Proposition 1 we have

$$V_A(\mu^D) \geq V_A(\mu) \text{ for all } \mu \in \mathcal{M}_R^D, \text{ in particular, for } \mu = \mu^O,$$

and

$$V_A(\mu^O) \geq V_A(\mu) \text{ for all } \mu \in \mathcal{M}_R^O, \text{ in particular, for } \mu = \mu^D.$$

It follows that $V_A(\mu^D) = V_A(\mu^O)$.

Next, μ^D is unimprovable by disclosure for party R, so

$$V_R(\mu^D) \geq V_R(\mu) \text{ for all } \mu \succeq \mu^D, \text{ in particular, for } \mu = \mu^O.$$

Also, μ^O is unimprovable by obfuscation for party R, so

$$V_R(\mu^O) \geq V_R(\mu) \text{ for all } \mu \preceq \mu^O, \text{ in particular, for } \mu = \mu^D.$$

It follows that $V_R(\mu^D) = V_R(\mu^O)$. Thus, we have reached a contradiction with the assumption that $\big(V_A(\mu^D), V_R(\mu^D)\big) \neq \big(V_A(\mu^O), V_R(\mu^O)\big)$.

Proof of Proposition 2

Denote full disclosure by μ^{FD} and no disclosure by μ^{ND}.

Part (i). Let party A choose μ^{FD}, so party R has no feasible deviation. To verify that this is an equilibrium, consider a potential deviation μ of party A. Applying Observation 1, we restrict attention to $\mu \in \mathcal{M}_R^D$. In particular, party R weakly prefers μ to μ^{FD}. But because μ^{FD} is Pareto undominated, it must be the case that party A weakly prefers μ^{FD} to μ, so μ is not a profitable deviation.

Part (ii). If μ^{ND} is Pareto dominant, then first $\mu^{ND} \in \mathcal{M}_R^D$, and second, μ^{ND} is in $\arg\max_{\mu \in \mathcal{M}_R^D} V_A(\mu)$. So by Observation 1, μ^{ND} is an equilibrium outcome.

Conversely, suppose that μ^{ND} is an equilibrium outcome. Then it must be the case that party R weakly prefers μ^{ND} to every outcome in $\mathcal{M}$. For a generic u_R, the outcome μ^{ND} is the unique most preferred outcome for party R, so

$$V_R(\mu^{ND}) > V_R(\mu) \text{ for all } \mu \succ \mu^{ND}, \tag{C1}$$

that is, party R is strictly worse of by any disclosure.

Now, suppose by contradiction that there exists an outcome μ that is strictly preferred to μ^{ND} by party A. Consider a small enough $\varepsilon > 0$ and an outcome $\tilde{\mu}_\varepsilon$ obtained by producing each message in the support of μ independently from the state with probability $1 - \varepsilon$ and

according to μ with probability ε. That is, $\tilde{\mu}_\varepsilon$ is in the ε-neighborhood of μ^{ND}, as it produces informative messages, but each of these messages is very close to being uninformative. By (C1) and the continuity of $V_R(\mu)$ w.r.t. μ,

$$V_R(\tilde{\mu}_\varepsilon) > V_R(\mu) \text{ for all } \mu \succ \tilde{\mu}_\varepsilon,$$

Thus, $\tilde{\mu}_\varepsilon \in \mathcal{M}_R^D$, so party R has no incentive to disclose more information. But by construction of $\tilde{\mu}_\varepsilon$, party A strictly prefers $\tilde{\mu}_\varepsilon$ to μ^{ND}. This is a contradiction to the assumption that μ^{ND} is an equilibrium outcome. ■

Proof of Lemma 1

We prove part (i) (the proof of part (ii) is symmetric). By (A_1), density g is continuously differentiable and strictly log-concave, so $g'(x)/g(x)$ is well defined, in particular, $g > 0$. Also, $u'_A > 0$ by assumption. Thus, by (1) we have

$$\begin{aligned} V''_A(x) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2} u_A(G(x)) = u''_A(G(x))\,(g(x))^2 + u'_A(G(x))g'(x) \\ &= u'_A(G(x))\,(g(x))^2 \left(\frac{u''_A(G(x))}{u'_A(G(x))} + \frac{g'(x)}{(g(x))^2} \right). \end{aligned} \tag{D1}$$

First, $u'_A(G(x))\,(g(x))^2 > 0$. Second, $G(x)$ is increasing, and $u''_A(y)/u'_A(y)$ is decreasing by (A_2), so $u''_A(G(x))/u'_A(G(x))$ is decreasing. Lastly, because $\ln g(x)$ is strictly concave, it follows that $g''(x)g(x) < (g'(x))^2$. Therefore,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x} \left(\frac{g'(x)}{(g(x))^2} \right) &= \frac{g''(x)(g(x))^2 - 2g(x)(g'(x))^2}{(g(x))^4} \\ &< \frac{(g'(x))^2 g(x) - 2g(x)(g'(x))^2}{(g(x))^4} = -\frac{(g'(x))^2}{(g(x))^3} \leq 0. \end{aligned}$$

Thus, g'/g^2 is strictly decreasing. We have proved that $V''_A(x)$ crosses the horizontal axis at most once and from above, which implies the statement of part (i). ■

Proof of Theorem 1

Suppose that $u'_A(q)/u'_R(1-q)$ is constant. Then (3) holds, implying that both full disclosure and no disclosure are Pareto undominated. Thus, by Propositions 2(i) and 3(i), full disclosure is an equilibrium outcome of the sequential disclosure game, and no disclosure is an equilibrium outcome of the sequential obfuscation game. Moreover, the parties are indifferent between equilibrium outcomes, that is,

$$(V_A(\mu'), V_R(\mu')) = (V_A(\mu''), V_R(\mu'')) \text{ for any equilibrium outcomes } \mu' \text{ and } \mu''. \tag{E1}$$

Suppose in addition that (A_1) and (A_2) hold. In sequential disclosure, the uniqueness of equilibrium outcome follows from Theorem 3. We now prove that in sequential obfuscation,

no disclosure, is the unique equilibrium outcome.

Let x_0 be the prior mean state, and let μ^{ND} be the no disclosure outcome, so μ^{ND} induces the posterior mean state x_0 with certainty. By contradiction, suppose that there is another equilibrium outcome $\mu^* \neq \mu^{ND}$. Then party R cannot improve upon μ^* by obfuscation, that is, $\mu^* \in \mathcal{M}_R^O$. We can interpret party R's decision problem as a standard Bayesian persuasion problem of a single sender, party R, with the distribution of the state given by μ^*, such that the optimal choice of party R is full disclosure of the state. As known in the literature (e.g., Kolotilin, 2018), for full disclosure to be optimal, it has to be the case that the sender cannot benefit by pooling (or partially pooling) any pair of states in the support. Formally, this can be expressed as follows. Let X^{μ^*} be the convex hull of the support of μ^*, so X^{μ^*} is the interval of posterior mean states that, potentially, can be induced by garbling of μ^*. Let $\bar{V}_R^{\mu^*}$ be the *augmented utility function* of party R given by

$$\bar{V}_R^{\mu^*}(x) = V_R(x) \text{ for each } x \in \operatorname{supp}(\mu^*), \tag{E2}$$

and $\bar{V}_R^{\mu^*}(x)$ is linearly extended on $X^{\mu^*} \backslash \operatorname{supp}(\mu^*)$. The condition that party R cannot benefit by pooling (or partial pooling) any pair of states in the support of μ^* is expressed as

$$\bar{V}_R^{\mu^*} \text{ is convex on } X^{\mu^*}, \text{ and } \bar{V}_R^{\mu^*}(x) \geq V_R(x) \text{ for all } x \in X^{\mu^*}. \tag{E3}$$

We have

$$V_R(x_0) = V_R(\mu^{ND}) = V_R(\mu^*) = \bar{V}_R^{\mu^*}(\mu^*) \geq \bar{V}_R^{\mu^*}(x_0) \geq V_R(x_0), \tag{E4}$$

where the first equality is because μ^{ND} induces prior mean x_0 with certainty, the second equality is by (E1) and the assumption that μ^* is an equilibrium outcome, the third equality is by (E2), the first inequality is Jensen's inequality due to the convexity of $\bar{V}_R^{\mu^*}$ and the fact that $\int x \mathrm{d}\mu^*(x) = x_0$, and the last inequality is by (E3) and the fact that x_0 is in X^{μ^*}.

From (E4), we conclude that the necessary condition for outcome μ^* to be an equilibrium outcome is that the graph of $(x, \bar{V}_R^{\mu^*}(x))_{x \in X^{\mu^*}}$ is a straight line that is weakly above V_R and is tangent to V_R at x_0, as illustrated in Fig. E1. However, by Lemma 1, when assumptions (A_1) and (A_2) hold, V_R has at most one inflexion point, so the above necessary condition cannot be satisfied. In other words, for any μ^* that is a mean-preserving spread of x_0, party R can strictly benefit by obfuscation. We reached a contradiction with the assumption that μ^* is an equilibrium outcome. ■

Proof of Theorem 2

Suppose that $u_A'(q)/u_R'(1-q)$ is decreasing. Let x_0 be the prior mean state, and let μ^{ND} be the no disclosure outcome, so μ^{ND} induces the posterior mean state x_0 with certainty. Consider a different utility function, $\tilde{u}_A$, for party A, given by

$$\tilde{u}_A(q) = u_A(G(x_0)) - \frac{u_A'(G(x_0))}{u_R'(1-G(x_0))}(u_R(1-q) - u_R(1-G(x_0))),$$

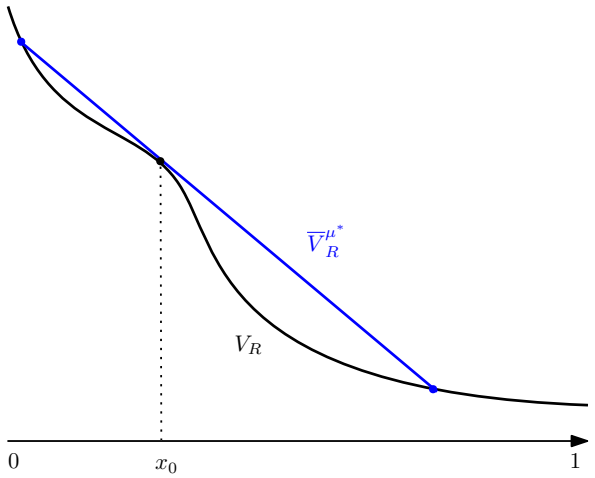

Figure E1. : No disclosure μ^{ND} (black dot), and μ^* with two-point support (blue dots) and the associated augmented utility $\bar{V}_R^{\mu^*}$.

and let $\tilde{V}_A(x) = \tilde{u}_A(G(x))$. Consider two problems:

$$(\mathrm{P_O}) \qquad \max_{\mu \in \mathcal{M}_R^O} V_A(\mu),$$

$$(\tilde{\mathrm{P}}_\mathrm{O}) \qquad \max_{\mu \in \mathcal{M}_R^O} \tilde{V}_A(\mu).$$

We will show that in $(\tilde{\mathrm{P}}_\mathrm{O})$ the maximum payoff for party A is attained at $\mu = \mu^{ND}$, that in $(\mathrm{P_O})$ the outcome μ^{ND} attains the same payoff for party A, and that $V_A(\mu) \leq \tilde{V}_A(\mu)$ for all μ. It will follow that μ^{ND} is a solution of $(\mathrm{P_O})$. Moreover, μ^{ND} is a unique solution of $(\mathrm{P_O})$ when either μ^{ND} is a unique solution of $(\tilde{\mathrm{P}}_\mathrm{O})$, or when $V_A(\mu) < \tilde{V}_A(\mu)$ for all $\mu \neq \mu^{ND}$.

To show the above, first, observe that $\tilde{u}_A$ and u_R satisfy CMUR, that is,

$$\frac{\tilde{u}_A'(q)}{u_R'(1-q)} = \frac{u_A'(G(x_0))}{u_R'(1-G(x_0))}$$

is constant in q. So, by Theorem 1, μ^{ND} is a solution of $(\tilde{\mathrm{P}}_\mathrm{O})$. Moreover, when (u_A, u_R, G) satisfy assumptions (A_1) and (A_2), function $\tilde{u}_A'$ is log-concave because u_A' is log-concave, so $(\tilde{u}_A, u_R, G)$ also satisfy assumptions (A_1) and (A_2). In this case, by Theorem 1, μ^{ND} is a unique solution of $(\tilde{\mathrm{P}}_\mathrm{O})$.

Second, because μ^{ND} induces x_0 with certainty, observe that

$$V_A(\mu^{ND}) = u_A(G(x_0)) = \tilde{u}_A(G(x_0)) = \tilde{V}_A(\mu^{ND}),$$

so the maximal payoff under $(\tilde{\mathrm{P}}_\mathrm{O})$ is attainable under $(\mathrm{P_O})$ by μ^{ND}.

Third, we show that $V_A(\mu) \leq \tilde{V}_A(\mu)$ for all $\mu \in \mathcal{M}$. It suffices to show that $u_A(q) \leq \tilde{u}_A(q)$ for all $q \in [0,1]$. Let $y = u_R(1-q)$ and let $y_0 = u_R(1-G(x_0))$. Then, substituting

$q = 1 - u_R^{-1}(y)$ and $G(x_0) = 1 - u_R^{-1}(y_0)$, we need to show that the expression

$$\tilde{u}_A(q) - u_A(q) = u_A(G(x_0)) - \frac{u'_A(G(x_0))}{u'_R(1-G(x_0))}(u_R(1-q) - u_R(1-G(x_0))) - u_A(q)$$

$$= u_A(1-u_R^{-1}(y_0)) - \frac{u'_A(1-u_R^{-1}(y_0))}{u'_R(u_R^{-1}(y_0))}(y-y_0) - u_A(1-u_R^{-1}(y)) \tag{F1}$$

is nonnegative for all $y \in [u_R(0), u_R(1)]$. Clearly, expression (F1) evaluated at $y = y_0$ is equal to zero, and its derivative w.r.t. y, which is given by

$$-\frac{u'_A(1-u_R^{-1}(y_0))}{u'_R(u_R^{-1}(y_0))} + \frac{u'_A(1-u_R^{-1}(y))}{u'_R(u_R^{-1}(y))}$$

is also equal to zero when evaluated at $y = y_0$. Moreover, by (4), expression (F1) is convex in y. We thus obtain that (F1) is nonnegative for all $y \in [u_R(0), u_R(1)]$.

Lastly, if $u'_A(q)/u'_R(1-q)$ is strictly decreasing, then expression (F1) is strictly convex in y, and thus (F1) is strictly positive for all $y \in [u_R(0), u_R(1)]\backslash\{y_0\}$. For this case, we conclude that $V_A(\mu^{ND}) = \tilde{V}_A(\mu^{ND})$, and $V_A(\mu) < \tilde{V}_A(\mu)$ for all $\mu \in \mathcal{M}\backslash\{\mu^{ND}\}$. ■

Proof of Theorem 3

Suppose that $u'_A(q)/u'_R(1-q)$ is increasing and assumptions (A_1) and (A_2) are satisfied. We will show, given any message m and an induced posterior distribution F_m of the state conditional on m, if

(i) F_m is nondegenerate, i.e., its support is nonsingleton, and

(ii) party R cannot benefit by revealing any information,

then party A strictly prefers to reveal the state.

This statement implies that full disclosure is a unique equilibrium outcome of sequential disclosure. This is because any deviation of party A from full disclosure must generate a message that is sent with a positive probability and leads to a nondegenerate posterior distribution of the state, and, conditional on this message, party R cannot benefit by revealing more information. But, as the above statement says, such a deviation cannot be profitable for party A. Conversely, if an outcome μ is not full disclosure and it is unimprovable by disclosure for party R, then there are messages sent with a positive probability that lead to nondegenerate posteriors, where party A has strictly profitable deviations.

To show the above statement, we prove two auxiliary lemmas.

First, recall that, by Lemma 1, under the assumptions (A_1) and (A_2), V_A is S-shaped with inflexion point τ_A, and V_R is inverted S-shaped with inflexion point τ_R. We show that if we assume IMUR, then the intervals where V_A and V_R are convex overlap, that is, the inflexion points satisfy $\tau_R \leq \tau_A$, as illustrated in Fig. 1.

LEMMA 2: *Suppose that $u'_A(q)/u'_R(1-q)$ is increasing and assumptions (A_1) and (A_2) are satisfied. Then $\tau_R \leq \tau_A$.*

PROOF:

Because $u'_A > 0$ and $u'_R > 0$ by assumption, the condition of increasing $u'_A(q)/u'_R(1-q)$ can be expressed as

$$\text{(G1)} \qquad \frac{u''_A(q)}{u'_A(q)} + \frac{u''_R(1-q)}{u'_R(1-q)} \geq 0 \quad \text{for all } q \in [0,1].$$

By Lemma 1, $V_A(x) = u_A(G(x))$ is strictly S-shaped with the inflexion point τ_A, i.e., $x < (>)\, \tau_A$ if and only if $V''_A(x) > (<)\, 0$. Because $u'_A > 0$, and the log-concavity of g implies that $g > 0$, it follows from (D1) that

$$\text{(G2)} \qquad x < (>)\, \tau_A \iff V''_A(x) > (<)\, 0 \iff \frac{u''_A(G(x))}{u'_A(G(x))} + \frac{g'(x)}{(g(x))^2} > (<)\, 0.$$

Also by Lemma 1, $V_R(x) = u_R(1 - G(x))$ is strictly inverted S-shaped with the inflexion point τ_R. By the symmetric argument we obtain

$$\text{(G3)} \qquad x < (>)\, \tau_R \iff V''_R(x) < (>)\, 0 \iff \frac{u''_R(1-G(x))}{u'_R(1-G(x))} - \frac{g'(x)}{(g(x))^2} < (>)\, 0.$$

By contradiction, suppose that $\tau_A < \tau_R$. Adding up inequalities (G2) and (G3) and considering x that satisfies $\tau_A < x < \tau_R$ we obtain

$$\frac{u''_A(G(x))}{u'_A(G(x))} + \frac{u''_R(1-G(x))}{u'_R(1-G(x))} < 0,$$

which contradicts (G1) with $q = G(x)$. We thus conclude that $\tau_A \geq \tau_R$.

Next, we provide a necessary and sufficient condition for party R to have no incentive to disclose information about the state. This is a direct adaptation of the analogous condition in the literature on Bayesian persuasion (e.g., Kolotilin, Mylovanov and Zapechelnyuk, 2022).

Fix a message m induced by a disclosure strategy of party A. Let F_m be a posterior probability distribution of the state conditional on m. Let $[a_m, b_m]$ be the closure of the convex hull of the support of F_m. Let x_m be the mean state under F_m, so $x_m = \int_{a_m}^{b_m} x \mathrm{d}F_m(x)$.

LEMMA 3: *Distribution F_m is unimprovable by disclosure for party R if and only if*

$$\text{(G4)} \qquad V_R(x_m) + V'_R(x_m)(x - x_m) \geq V_R(x) \quad \textit{for all } x \in [a_m, b_m].$$

PROOF:

Let us interpret the problem that party R faces after observing message m as a standard Bayesian persuasion problem of a single sender, party R, with the prior F_m about the state. Because V_R is strictly inverted S-shaped by Lemma 1, it follows from Kolotilin, Mylovanov and Zapechelnyuk (2022) (after an appropriate normalization) that there is a

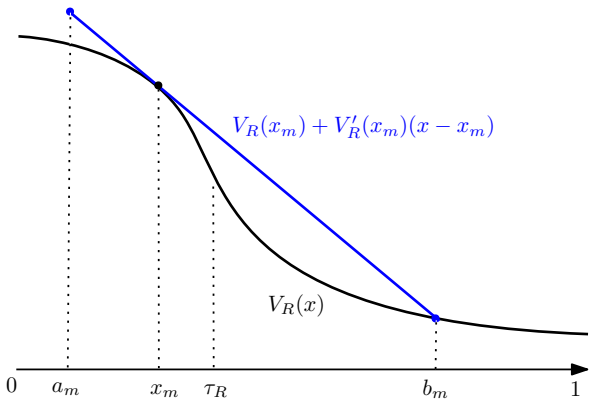

Figure G1. : Optimality of no disclosure for party R when the state is distributed on interval $[a_m, b_m]$ with mean x_m.

cutoff $\tilde{x} \in [a_m, b_m]$ such that party R optimally reveals the states in the interval $(\tilde{x}, b_m]$ and pools the states in the interval $[a_m, \tilde{x}]$. Moreover, the optimal cutoff $\tilde{x}$ coincides with the upper bound b_m (so all states in $[a_m, b_m]$ are pooled) if and only if (G4) holds, as illustrated in Fig. G1.

We now return to the proof of Theorem 3. Consider a nondegenerate posterior distribution of the state F_m over interval $[a_m, b_m]$ with mean x_m such that party R cannot benefit by disclosure, so condition (G4) of Lemma 3 holds. Observe that x_m must be in the interval where V_A is concave, that is,

$$x_m < \tau_A. \tag{G5}$$

Otherwise, if $x_m \geq \tau_A$, then, by Lemma 2 we have $x_m \geq \tau_R$, so $V_R(x)$ is strictly convex in the neighborhood of x_m. But then condition (G4) of Lemma 3 cannot be satisfied for nondegenerate F_m.

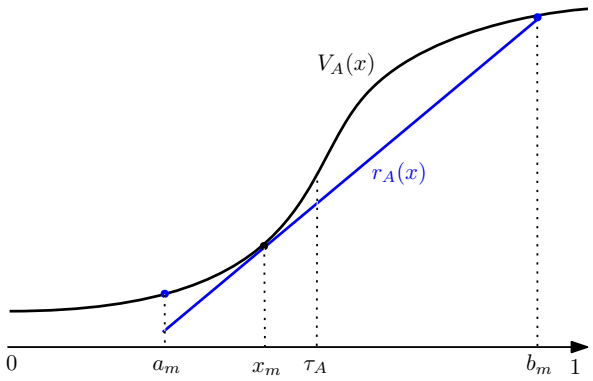

Figure G2. : Optimality of full disclosure for party A when the state is distributed on interval $[a_m, b_m]$ with mean x_m.

Next, let $r_A(x)$ be the tangency line to $V_A(x)$ at x_m,

$$r_A(x) = V_A(x) - V_A(x_m) - V_A'(x_m)(x - x_m),$$

as shown in Fig. G2. We need to show that

$$V_A(x) > r_A(x) \text{ for all } x \in [a_m, b_m)\backslash\{x_m\}. \tag{G6}$$

Once we have shown (G6), it will follow immediately that

$$\int_{a_m}^{b_m} V_A(x)\mathrm{d}F_m(x) > V_A(x_m),$$

that is, party A strictly prefers to fully reveal the state. This will complete the proof of Theorem 3.

Let us prove (G6). Because V_A is continuous and strictly S-shaped by assumptions (A_1) and (A_2) and Lemma 1, and because $x_m < \tau_A$ by (G5), it is apparent from Fig. G2 that for (G6) to hold, it suffices to prove that $V(a_m) > r_A(a_m)$ and $V(b_m) \geq r_A(b_m)$. Because V_A is strictly concave on $[a_m, \tau_A)$ and $x_m \in (a_0, \tau_A)$ we obtain

$$V_A(a_m) - r_A(a_m) = V_A(a_m) - V_A(x_m) - V_A'(x_m)(a_m - x_m) > 0.$$

It remains to show that $V(b_m) \geq r_A(b_m)$. Recall that (G4) is assumed to hold. Let $y_m = V_R(x_m)$ and $y_b = V_R(b_m)$, so $x_m = V_R^{-1}(y_m)$ and $b_m = V_R^{-1}(y_b)$. Substituting these into (G4) with $x = b_m$ and rearranging the terms (taking into account that $V_R' < 0$) yields

$$V_R^{-1}(y_b) - V_R^{-1}(y_m) \leq \frac{y_b - y_m}{V_R'(V_R^{-1}(y_m))}. \tag{G7}$$

Next, we have

$$\begin{aligned} V(b_m) - r_A(b_m) &= V_A(b_m) - V_A(x_m) - V_A'(x_m)(b_m - x_m) \\ &= V_A(V_R^{-1}(y_b)) - V_A(V_R^{-1}(y_m)) - V_A'(V_R^{-1}(y_m))(V_R^{-1}(y_b) - V_R^{-1}(y_m)) \\ &\geq V_A(V_R^{-1}(y_b)) - V_A(V_R^{-1}(y_m)) - \frac{V_A'(V_R^{-1}(y_m))}{V_R'(V_R^{-1}(y_m))}(y_b - y_m) \\ &\geq 0, \end{aligned}$$

where the second line is by the substitution $x_m = V_R^{-1}(y_m)$ and $b_m = V_R^{-1}(y_b)$, the third line is by $V_A' > 0$ and inequality (G7), and the last line is because

$$V_A(V_R^{-1}(y)) = u_A(1 - u_R^{-1}(y))$$

is convex in y by (5). This completes the proof of (G6). ■