Theoretical Economics 9 (2014), 99–136 1555-7561/20140099

Persuasion and dynamic communication

Itai Sher
Department of Economics, University of Minnesota

A speaker attempts to persuade a listener to accept a request by presenting evidence. A *persuasion rule* specifies what evidence is persuasive. This paper compares *static* and *dynamic* rules. We present a single linear program (i) whose solution corresponds to the listener's optimal dynamic rule and (ii) whose solution with additional integer constraints corresponds to the optimal static rule. We present a condition—*foresight*—under which the optimal persuasion problem reduces to the classical maximum flow problem. This has various qualitative consequences, including the coincidence of optimal dynamic and static persuasion rules, elimination of the need for randomization, and symmetry of optimal static rules.

Keywords. Communication, optimal persuasion rules, credibility, commitment, evidence, maximum flow problem.

JEL classification. C61, D82, D83.

1. Introduction

This paper compares two modes of persuasion: *static* persuasion and *dynamic* persuasion. In both scenarios, a speaker presents evidence to a listener to persuade the listener to take the speaker's preferred action. Static persuasion is one shot: the speaker simply presents evidence and the listener responds with a decision. Dynamic persuasion involves back-and-forth communication. The speaker makes an opening argument via a cheap talk claim. The listener then requests some of the evidence that the speaker claims to have. The speaker responds by presenting either the requested evidence or some other evidence. Finally, the listener makes his decision. Through the comparison of static and dynamic persuasion, we aim to attain an enriched understanding (i) of the determination of what evidence will be persuasive and (ii) of when and how dynamic back-and-forth cheap talk communication complements evidence production in strategic persuasion.

Itai Sher: isher@umn.edu

A previous version of this paper circulated under the title "Persuasion and limited communication." I am very grateful to Eddie Dekel for many valuable comments and advice. I would also like to thank David Austen-Smith, Marciano Siniscalchi, Asher Wolinsky, David Rahman, Ariel Rubinstein, Jesse Bull, Elchanan Ben-Porath, Melissa Koenig, Peter Sher, Tomasz Strzalecki, Evsen Turkay, and Jan Werner. I am grateful to Rakesh Vohra for suggesting the relation of the persuasion problem to the set cover problem. I thank Bart Lipman and two anonymous referees for their helpful comments. Shihui Ma, Nyalleng Moorosi, and Benjamin Segal provided valuable research assistance. I thank Kemal Badur for helping to construct the figures for this paper. A grant-in-aid and single semester leave from the University of Minnesota are gratefully acknowledged. All errors are mine.

DOI: 10.3982/TE683

Dynamic persuasion may outperform static persuasion when it is infeasible for the speaker to present all evidence due to costs associated with presenting or processing evidence. Then the opening argument guides the subsequent back-and-forth discussion. For example, in a time-constrained interview, a job candidate must strategically decide which qualifications to emphasize. If she emphasizes her leadership capacity, the employer may ask her how she would lead certain projects and to provide evidence of her past leadership. If she emphasizes her analytical ability, the employer may give her a puzzle to solve. The job seeker may be afraid to make false claims for fear of being asked for supporting evidence that she cannot supply. It often benefits the listener to randomize her response to the speaker's claims, because if the speaker can predict what follow-up questions will be asked, she may lie about everything else.

This paper provides a unified framework for thinking about both static and dynamic persuasion, and the relationship between the two. We focus on mechanisms—called *persuasion rules*—that are optimal for the listener. Our main contributions are the following:

1. We characterize the relation between optimal dynamic and static persuasion by presenting a single linear program (i) whose solution corresponds to the optimal dynamic persuasion rule and (ii) whose solution with additional integer constraints corresponds to the optimal static persuasion rule.

2. We present a condition with a natural interpretation—which we call *foresight*—under which our linear program simplifies to the classical maximum flow problem. This leads to various qualitative consequences, including the coincidence of optimal dynamic and static persuasion rules, elimination of the need for randomization, and symmetry of optimal static rules.

The remainder of the introduction elaborates on these two contributions. We discuss the two results in turn.

1.1 *Characterization of optimal persuasion*

Our first main contribution is the exact simultaneous characterization of optimal static and dynamic persuasion: Theorem 1 presents a linear programming characterization of optimal dynamic persuasion, and Theorem 2 shows that by adding an integer constraint, one obtains an integer programming characterization of optimal static persuasion. This result serves as a foundation for all of our other results, including the credibility result described below and the stronger characterization under the additional foresight assumption described in Section 1.2. Being an exact characterization, these programs should serve as the starting point for future studies of the persuasion problem or variants thereof.

Our characterization has an intuitive interpretation: it provides a precise sense in which static persuasion equals dynamic persuasion minus randomization. While static rules allow for randomization of the listener's decision, randomization is only beneficial as part of a back-and-forth conversation.[1] Conversely, such back-and-forth communication can be useful only if the listener's questions are unpredictable. If a question is predictable, it need not be asked; the speaker will know at the outset that she must answer it. So a necessary and sufficient condition for dynamic communication to improve on static communication is that the listener can benefit by randomizing his questions.

Glazer and Rubinstein (2004, 2006) initiated the study of optimal persuasion rules.[2] Our results shed interesting light on the main results of Glazer and Rubinstein. Glazer and Rubinstein use a constraint called the L-principle to analyze persuasion. We characterize the circumstances in which the L-principle captures all of the incentive constraints involved in optimal persuasion, and show that the L-principle is not, in general, sufficient for analyzing optimal dynamic persuasion. Our more general analysis then allows us to extend an important result of Glazer and Rubinstein to a broader class of persuasion problems. In particular, the optimal persuasion problem assumes listener commitment, but we show that in the dynamic persuasion problem, commitment has no value for the listener. In other words, there is a sequential equilibrium of the dynamic persuasion game in which the listener cannot commit to his strategy up front and that implements the same outcome as the optimal rule. This *credibility result* generalizes a result of Glazer and Rubinstein (2004) beyond the case to which the L-principle applies. The dual of the program in our result that characterizes optimal dynamic persuasion—Theorem 1—is used to derive the speaker's strategy in the credible implementation of the optimal rule.

1.2 *Foresight and Max Flow*

Our second main contribution is a reduction of the optimal persuasion problem to the classical maximum flow problem (Max Flow) under an intuitive and easily verifiable condition that we call *foresight* (Theorem 6). Max Flow is a very extensively studied optimization problem from the field of combinatorial optimization (see, for example, Ahuja et al. 1993). Our reduction allows us to provide a condition under which optimal static and dynamic rules coincide, randomization is not needed for optimal persuasion, and optimal static rules are symmetric in the sense that similar evidence is treated similarly. The reduction also generates monotone comparative statics for persuasion, which can be found in an earlier version of this paper (Sher 2008). A fundamental theorem on the maximum flow problem is the max-flow min-cut theorem (Ford and Fulkerson 1956), which is a special case of linear programming duality. In Section 5.2, we illustrate how the credibility result mentioned above (in Section 1.1) coincides with max-flow min-cut duality under foresight.

[1] Glazer and Rubinstein (2006) provide an informal discussion that makes a similar point about the role of randomization in the context of discussing a special case of the persuasion problem from Glazer and Rubinstein (2004) that we call $(1,2)$-verification environments below.

[2] There is a broader literature on persuasion games (Milgrom and Roberts 1986, Shin 1994a, Glazer and Rubinstein 2001, Chen and Olszewski 2011, Dziuda 2011). A different strand of the literature studied dynamic communication in pure cheap talk frameworks (Krishna and Morgan 2004, Aumann and Hart 2003). The study of cheap talk was pioneered by Crawford and Sobel (1982). The optimal mechanism in a cheap talk framework was studied by Goltsman et al. (2009).

Foresight has a natural interpretation. The optimal persuasion problem contains two types of states: *accept states* and *reject states*. The listener would like to be persuaded by the speaker only in accept states. Define evidence to be *genuine* whenever it is presented in an accept state and *forged* whenever it is presented in a reject state. Foresight means that at every accept state x, there is evidence m_x that is maximally difficult to forge (among evidence available at x) in the sense that m_x is available least often in reject states (measured by inclusion). We call the property *foresight* because in the dynamic communication protocol, the speaker should be able to foresee that if she claims the state is x (an accept state), the listener would require evidence m_x as support (if the claim x has any chance of being persuasive). This contrasts with the general case in which the speaker faces a lottery over evidence requests.

In the context of the persuasion environment studied here, foresight generalizes the concept of normality from the literature on mechanism design with evidence (Bull and Watson 2007). Closely related properties are studied by Green and Laffont (1986) and Lipman and Seppi (1995). Remark 1 and the discussion following Corollary 1 in Section 5 highlight the ways in which foresight is weaker than normality. In the conclusion (Section 7), we explain how the basic idea underlying foresight can be extended to more general mechanism design problems with evidence.

The organization of the paper is as follows. Section 2 presents the model. Section 3 presents the linear programming formulation of optimal persuasion. Section 4 uses this formulation to attain a new credibility result. Section 5 studies the consequences of foresight. Section 6 briefly provides some consequences of the preceding analysis for the complexity of optimal rules. Section 7 concludes. The Appendix contains many of the proofs.

2. Model

Section 2.1 presents the persuasion environment. Section 2.2 presents two modes of persuasion: *static* and *dynamic* persuasion (Figures 1 and 2). The model of static persuasion is the model of Glazer and Rubinstein (2006) and the model of dynamic persuasion generalizes Glazer and Rubinstein (2004). Section 2.3 establishes the robustness of the model of dynamic persuasion.

2.1 *Environment*

Let X be a finite set of states. $A \subseteq X$ is the set of *accept states* and $R = X \setminus A$ is the set of *reject states*. For $x \in X$, $p_x > 0$ is the probability of x. Let $p = (p_x)_{x \in X}$.

A speaker wishes to persuade a listener to accept a request. For example, the speaker may be attempting to persuade the listener to hire her. Alternatively, the speaker may be a lobbyist attempting to persuade the listener, a legislator, to support some legislation. The speaker always wants to be accepted regardless of the state. Ideally, the listener would like to accept the speaker when the state is in A and to reject the speaker when the state is in R. If the listener is uncertain about the state, he aims to minimize the *error probability*, that is, the probability that he accepts in R or rejects in A.

1. The true state x is realized according to probability distribution p and is known only to the speaker.
2. The speaker sends a message $m \in \sigma(x)$.
3. The listener accepts the speaker's request with probability $f(m)$.

FIGURE 1. Static communication protocol.

For every state $x \in X$, there is a finite nonempty set $\sigma(x)$ of statements that are available at x. We refer to σ is the *message correspondence*. Let $M = \bigcup_{x \in X} \sigma(x)$. There exists $m^0 \in M$ such that for all $x \in X$, $m^0 \in \sigma(x)$. The message m^0 is called the *vacuous message* because it proves nothing about the state. The message m^0 simplifies proofs but is inessential to our results. We generally omit m^0 in examples.

The message correspondence models a communication constraint. As different messages are available at different states, the messages amount to evidence. Imagine, for example, that an unqualified person interviews for a computer programmer position. The interviewer asks how the interviewee would program a certain task. The interviewee—being unqualified—is unable to answer. Were she qualified, she would be able to answer.

An important aspect of the constraint encoded by σ is that the speaker *may only present one message* $m \in \sigma(x)$. This may be a time constraint. To see that this is essentially without loss of generality, define

$$\sigma^n(x) := \{S \subseteq \sigma(x) : |S| \leq n\}.$$

If we wanted to relax the time constraint and let the speaker present n messages instead of just one, we could simply replace σ by the message correspondence σ^n.

2.2 *Two modes of persuasion*

We introduce static persuasion and dynamic persuasion. The static timing is given by the static communication protocol (Figure 1). The listener's strategy is determined by a function $f : M \to [0, 1]$, called a *static persuasion rule*; F is the set of such static rules.

While static persuasion is restrictive, Proposition 1 below shows that the dynamic persuasion protocol (Figure 2) is canonical: The listener can minimize his error probability among all *possible* mechanisms via the dynamic communication protocol. A listener's strategy specifies the probability that he requests various messages (step 3(a) of Figure 2) and that he rejects the speaker's request outright (step 3(b) of Figure 2). (The listener's behavior is fixed at step 5 of Figure 2.) Formally, the listener's strategy is a function $g : X \times M \to [0, 1]$ that satisfies

$$\forall \hat{x} \in X, \quad \sum_{m \in M} g(\hat{x}, m) \leq 1 \tag{1}$$

$$\forall \hat{x} \in X, \forall m \notin \sigma(\hat{x}), \quad g(\hat{x}, m) = 0. \tag{2}$$

1. The true state x is realized according to probability distribution p and is known only to the speaker.
2. The speaker makes a cheap talk claim that the state is $\hat{x}$.
3. (a) For each $m \in \sigma(\hat{x})$, the listener requests that the speaker present message m with probability $g(\hat{x}, m)$ and the protocol proceeds to step 4.

   (b) With probability $1 - \sum_{m\in\sigma(\hat{x})} g(\hat{x}, m)$, the listener rejects the speaker's request and the communication protocol ends without further communication.
4. The speaker presents some message $m' \in \sigma(x)$.
5. (a) If the speaker fails to present the requested message (i.e., $m' \neq m$), the listener rejects the speaker's request.

   (b) If the speaker presents the requested message (i.e., $m' = m$), the listener accepts the speaker's request.

FIGURE 2. Dynamic communication protocol.

Equation (2) says that the listener never requests a message the speaker claims is unavailable. Inequality (1) allows us to interpret $g(\hat{x}, m)$ as a probability. A function g that satisfies (1) and (2) is a *dynamic persuasion rule*, and G is the set of such dynamic rules.

For both protocols, we assume the listener commits to a persuasion rule and the speaker best replies. (We relax this commitment assumption in Section 4.) Any persuasion rule then induces an acceptance probability at every state. For a static persuasion rule f, the acceptance probability is $\alpha(f, x) := \max_{m\in\sigma(x)} f(m)$ because the speaker will simply select the *available* message that maximizes the acceptance probability. For a dynamic persuasion rule g, the acceptance probability is

$$\alpha(g, x) := \max_{z\in X} \sum_{m\in\sigma(z)\cap\sigma(x)} g(z, m).$$

To understand this, observe that if the speaker claims the state is z when the state is really x, the listener will request only messages m that belong to $\sigma(z)$ and the speaker will only succeed in presenting the required messages if they belong to $\sigma(x)$ as well. At state x, the speaker will choose the cheap talk report z that maximizes the probability of acceptance.

Given a persuasion rule h (static or dynamic), the induced *error probability at state* x is then given by $\mu_x(h) := 1 - \alpha(h, x)$ if $x \in A$ and $\mu_x(h) := \alpha(h, x)$ if $x \in R$. The *total error probability* is then $\mu(h) := \sum_{x\in X} p_x\mu_x(h)$. The *static persuasion problem* is to select $f \in F$ to minimize $\mu(f)$ and the *dynamic persuasion problem* is to select $g \in G$ to minimize $\mu(g)$. Nothing essential would change if we introduce a state dependent error cost $\ell_x > 0$. Then the listener would minimize $\sum_{x\in X} \ell_x p_x\mu_x(h)$. One could derive analogs of all our results by substituting $\ell_x p_x$ for p_x.

2.3 *The dynamic protocol is canonical*

We now establish that nothing is gained by considering mechanisms more general than the dynamic communication protocol. Consider an arbitrary finite two-player extensive form game with imperfect information and perfect recall Γ. (See Osborne and Rubinstein 1994 for precise definitions.) The two players are the speaker and listener. The game Γ has no moves of nature, although allowing moves of nature does not alter Proposition 1.[3] Every terminal history is labeled either *accept* or *reject*. There are two types of histories among those at which the speaker moves: *evidentiary* and *ordinary*. At evidentiary histories, the set of available actions is M; no restrictions are placed on ordinary histories. Define $\Gamma_\sigma(x)$ to be the game formed by deleting all actions $m \in M \setminus \sigma(x)$ at each evidentiary history as well as the histories emanating from these deleted actions. So at any evidentiary history in $\Gamma_\sigma(x)$, $\sigma(x)$ is the set of available speaker actions. For any natural number n, write $\Gamma \in \mathbb{G}^n$ if each history in Γ has at most n evidentiary subhistories. (A *subhistory* of history h is an *initial* subsequence of h, possibly the empty sequence.) So if Γ belongs to $\mathbb{G}^n$, the speaker has at most n opportunities to present evidence at Γ. A *mechanism* is a pair (Γ, λ), where Γ is a game that meets the above specifications and λ is a behavioral strategy for the listener in Γ. Strategy λ induces a listener strategy in every game $\Gamma_\sigma(x)$ for $x \in X$. In particular, every information set I in $\Gamma_\sigma(x)$ corresponds to some information set I' in Γ at which the listener has observed the same sequence of events.[4] Strategy λ specifies that the listener will behave in I as he would in I'. Our formulation encodes the assumption that the listener commits to his strategy. We assume that at each state x, the speaker selects a best reply to λ in $\Gamma_\sigma(x)$.

Proposition 1. *The minimal error probability among all mechanisms* (Γ, λ) *with* $\Gamma \in \mathbb{G}^n$ *is attained by the optimal dynamic persuasion rule for message correspondence* σ^n.

Proposition 1 is closely related to Theorem 6 of Bull and Watson (2007), which presents a canonical mechanism (the *special three-stage mechanism*) for dynamic (weak) implementation. Due to the special structure of the persuasion problem, the dynamic communication protocol has more structure than the special three-stage mechanism: the option to reject the speaker outright (at step 3(b) of Figure 2) and the requirement to accept the speaker exactly if she presented the requested evidence are absent from the special three-stage mechanism (in step 5 of Figure 2). Indeed "acceptance" and "rejection" have no meaning in the more general framework of Bull and Watson (2007). Another difference is that whereas Bull and Watson's Theorem 6 applies to dynamic mechanisms where each terminal history contains exactly one evidentiary subhistory (per player), which is natural given their concerns, Proposition 1 shows how dynamic persuasion rules (weakly) improve upon arbitrary dynamic mechanisms that allow the

[3] Because Proposition 1 assumes that the listener can commit to her strategy, for the purposes of that proposition, any game with moves of nature Γ can be replicated by a game Γ', where these moves are replaced by moves of the listener at which the listener's strategy specifies that he randomizes as nature would have randomized in Γ.

[4] Note, however, that not every information set in Γ corresponds to an information set in $\Gamma_\sigma(x)$.

speaker to present evidence up to a fixed number of times n. Our result shows that given any time constraint, we can restrict attention to the dynamic communication protocol, using message correspondence σ if one round of evidence transmission is allowed, and, more generally, simply reinterpreting the message correspondence as σ^n if n rounds of evidence transmission are allowed.

3. Optimal persuasion rules

3.1 *Programs for optimal rules*

We now present optimization problems whose solutions are optimal static and dynamic persuasion rules. A feasible solution is of the form

$$(\mu, \beta) = (\{\mu_z\}_{z\in X}, \{\beta(x, m)\}_{x\in A, m\in\sigma(x)}).$$

Program P0. Minimize

$$\sum_{z\in X} \mu_z p_z \tag{3}$$

subject to

$$1 - \mu_x = \sum_{m\in\sigma(x)} \beta(x, m) \quad \forall x \in A \tag{4}$$

$$\mu_y \geq \sum_{m\in\sigma(x)\cap\sigma(y)} \beta(x, m) \quad \forall x \in A, \forall y \in R \tag{5}$$

$$\mu_z \geq 0 \quad \forall z \in X \tag{6}$$

$$\beta(x, m) \geq 0 \quad \forall x \in A, \forall m \in \sigma(x). \tag{7}$$

For each $z \in X$, μ_z is the probability of error at state z. The objective is to minimize the total error probability. We may interpret $\beta(x, m)$ as $g(x, m)$, the probability that the dynamic persuasion rule requests the message m if the speaker claims that the state is x. Whereas $g(x, m)$ is defined for all $x \in X$, $\beta(x, m)$ is only defined for states $x \in A$ (and $m \in \sigma(x)$). Intuitively, the program restricts attention to dynamic rules that always reject the speaker if she claims that the state is in R. Equation (4) can then be interpreted as saying that for any state $x \in A$, the error probability at x equals the probability that the rule would reject the speaker at state x *if she truthfully reported the state*. Inequality (5) says that for any state $y \in R$ and any state $x \in A$, the error probability at state y is at least the acceptance probability that the speaker would obtain at state y if she claimed that the state were x. Inequality (5) implies that at each state $y \in R$, the speaker is optimizing. But Program P0 contains no constraints that directly impose optimizing behavior at states $x \in A$; the program assumes via (4) that for each $x \in A$, the speaker reports truthfully, but there are no corresponding incentive constraints.

Theorem 1. *Let (μ^*, β^*) be an optimal solution to Program P0. Then g^* defined by*

$$g^*(x, m) = \begin{cases} \beta^*(x, m) & \text{if } x \in A, m \in \sigma(x) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

is an optimal dynamic persuasion rule. The rule g^* induces error probability μ_z^* at state z.

Program P0 characterizes optimal *dynamic* persuasion rules. Glazer and Rubinstein (2006) show that when restricting attention to *static* persuasion rules, there always exists an optimal static rule f^* that is *deterministic*, meaning that $f^*(m) \in \{0,1\}$ for all m (see their Proposition 1). Glazer and Rubinstein (2006) also argue informally—with respect to the special case of $(1,2)$-verification problems to be defined in Section 3.2—that dynamic persuasion rules without randomization reduce to static persuasion rules. We now establish the latter reduction argument for a wider class of persuasion environments, and in the process, specify the precise relationship between optimal static and optimal dynamic rules. To do this, we consider two ways to add integrality constraints to Program P0.

Program P1. Minimize (3) subject to (4)–(6) and

$$\beta(x,m) \in \{0,1\} \quad \forall x \in X, \forall m \in \sigma(x). \tag{9}$$

Program P2. Minimize (3) subject to (4), (5), (7), and

$$\mu_x \in \{0,1\} \quad \forall x \in X. \tag{10}$$

Theorem 2. *(i) Let (μ^*, β^*) be an optimal solution to Program P1. Then f^* defined by*

$$f^*(m) := \begin{cases} \max\{\beta^*(x,m) : x \in A\} & \text{if } m \in \bigcup_{x\in A}\sigma(x) \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

is an optimal static persuasion rule. The rule f^* induces error probability μ_z^* at state z.

(ii) Any optimal solution to Program P1 is an optimal solution to Program P2. For any optimal solution (μ^*, β^*) to Program P2, there exists β^{**} such that (μ^*, β^{**}) is an optimal solution to Program P1.

The theorem shows that adding integer constraints transforms Program P0 from a program for optimal dynamic persuasion to a program for optimal static persuasion. This can be done in two distinct ways: *either* by imposing a no-randomization constraint on the listener's requests (Program P1) *or* by imposing a no-randomization constraint on the state-by-state error probabilities (Program P2).[5] Thus we have a single formulation for both the static and the dynamic persuasion problem, which differs only insofar as we restrict our search to integral solutions.

This highlights the role of randomization. In static persuasion, randomization is permitted but inessential, as there always exists an optimal static persuasion rule that

[5]Given an optimal solution to Program P2, the proof of part (ii) of Theorem 2 shows how to derive an optimal solution to Program P1 and, hence, how to derive an optimal static persuasion rule via (11).

is deterministic. There, the listener can only randomize the accept/reject decision in response to a message. In contrast, in the dynamic persuasion problem, the listener randomly requests evidence in response to a cheap talk claim about the state. Theorem 2 implies that the elimination of *random* requests reduces dynamic persuasion to static persuasion.

Theorem 2 implies that optimal dynamic rules have weakly lower error probability than optimal static rules because the static problem is derived from the dynamic problem by adding constraints.[6] Example 1 shows that this inequality may sometimes be strict.

Example 1. Let $X := \{0,1\}^3$ and $A := \{x : \sum_{i=1}^{3} x_i \text{ is even}\}$. (We consider 0 to be even.) For $z = (z_1, z_2, z_3) \in X$, let $\sigma(z) := \{(z_i, i) : i = 1, 2, 3\}$. So $\sigma((0,1,0)) = \{(0,1), (1,2), (0,3)\}$. Message $(1,2)$ reveals that the second component of the state is 1 and message $(0,3)$ reveals that the third component is 0. Let $p_z = \frac{1}{8}$ for all $z \in X$. An optimal static persuasion rule is

$$f^*(m) = \begin{cases} 1 & \text{if } m = (0,i) \text{ for } i \in \{1,2,3\} \\ 0 & \text{otherwise.} \end{cases}$$

The rule f^* leads the speaker to be accepted at every state except $(1,1,1)$, inducing an error probability of $\frac{3}{8}$. In contrast, an optimal dynamic persuasion rule is

$$g^*(x,m) := \begin{cases} \frac{1}{3} & \text{if } x \in A \text{ and } m \in \sigma(x) \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

The rule g^* leads the speaker to be accepted with probability 1 at every state in A and to be accepted with probability $\frac{2}{3}$ in every state in R, inducing an error probability of $\frac{1}{3}$. Because $\frac{1}{3} < \frac{3}{8}$, the optimal dynamic persuasion rule g^* outperforms the optimal static persuasion rule f^*. ◊

3.2 *Insufficiency of the L-principle*

We compare our approach to that of Glazer and Rubinstein. An L is a pair (x, T) with $x \in A$, $T \subseteq R$, and $\sigma(x) \subseteq \bigcup_{y \in T} \sigma(y)$. An L, (x,T), is *minimal* if for all proper subsets T' of T, (x, T') is not an L. Glazer and Rubinstein introduce the *L-principle*

$$\sum_{z \in \{x\} \cup T} \mu_y \geq 1 \quad \text{for every minimal } L, (x,T). \tag{13}$$

To understand this, consider an L, (x,T), and observe that if a deterministic static rule f avoids a mistake at $x \in A$, f must accept some message $m \in \sigma(x)$. Then $\mu_x = 0$, so $\sum_{y \in T} \mu_y \geq 1$. Why? Because at least one $y \in T$ also has m. So f must accept the speaker at y, committing a mistake. Similar logic applies with randomization and even to dynamic rules. The following class of environments helps us to understand to scope of the L-principle.

[6] This is also a consequence of Proposition 1.

(k,n)-Verification Environment. Let k and n be natural numbers with $1 \leq k < n$. Let X_i, $i = 1, \dots, n$, be a collection of finite sets with $|X_i| \geq 2$ for all i and let the set of states X have the structure $X = X_1 \times X_2 \times \cdots \times X_n$. For any $J \subseteq \{1, \dots, n\}$ and $x = (x_1, \dots, x_n) \in X$, define $x_J := \{(x_j, j) : j \in J\}$ and let $\sigma(x) := \{x_J : |J| \leq k\}$. So, at each state, the speaker observes a vector $(x_1, \dots, x_n)$. The speaker can reveal (at most) k components of the vector. If the speaker reveals a component, she reveals both its value ($x_i \in X_i$) and its index ($i \in \{1, \dots, n\}$). The speaker is restricted to reveal at most k components, but it is feasible for her to reveal *any* k components.

Glazer and Rubinstein (2006) study *static* persuasion for general environments from Section 2.1, using Program GR1 below. Glazer and Rubinstein (2004) study *dynamic* persuasion, but only for $(1,2)$-verification environments,[7] using a special case of Program GR0:[8]

Program GR0. Minimize $\sum_{z \in X} p_z \mu_z$ subject to the L-principle (13) and

$$\mu_z \in [0,1] \quad \forall z \in X.$$

Program GR1. Minimize $\sum_{z \in X} p_z \mu_z$ subject to the L-principle (13) and

$$\mu_z \in \{0,1\} \quad \forall z \in X.$$

Theorem 3. *(i) (Glazer and Rubinstein 2004) Let $(\mu^*_x)_{x \in X}$ be an optimal solution to Program GR0 for some $(1,2)$-verification environment. Then there is an optimal dynamic persuasion rule g^* such that for all states $x \in X$, g^* induces error probability μ^*_x at x.*

(ii) (Glazer and Rubinstein 2006) Let $(\mu^*_x)_{x \in X}$ be a solution to Program GR1 for any persuasion problem. Then there is an optimal static persuasion rule f^* such that for all states $x \in X$, f^* leads to error probability μ^*_x at state x.

Unfortunately, it impossible to extend the first part of the result to general (k,n)-verification environments and, hence, also to general persuasion environments.

Theorem 4. *(i) Any optimal solution to Program GR0 gives a lower bound on the error probability at the optimal dynamic persuasion rule.*

[7] In Glazer and Rubinstein (2004), rather than the speaker presenting evidence, the listener verifies an aspect of the speaker's claim, but for $(1,2)$-verification environments, the two are mathematically equivalent.

[8] Glazer and Rubinstein (2004) discuss extending the L-principle to $(1,n)$-verification environments for $n > 2$ in a dynamic setting. They mention L's of the form $(x, \{y^i, y^j\})$ such that x differs from y^i exactly on component i and differs from y^j exactly on component j, where $i \neq j$, but neglect to mention other minimal L's in the sense of the more general definition they later presented in Glazer and Rubinstein (2006), such as those of the form $(x, \{y^i : i = 1, \dots, n\})$, where y^i agrees with x only on component i. The neglected minimal L's are all *valid* in the sense that they must be satisfied by any dynamic persuasion rule. Omitting these L's only makes the problem identified by Theorem 4 below worse, because that theorem shows that the L-principle imposes *too few* constraints in a dynamic setting.

(ii) Fix natural numbers n and k with $1 \leq k \leq n$. The following statements are equivalent:

(a) For all (k, n)-verification environments, the value of the corresponding Program GR0 is equal to the error probability at the optimal dynamic persuasion rule.

(b) The equation $n - k = 1$ holds.

Part (i) and its proof show that the L-principle is *valid* for general persuasion problems. In other words, the error probabilities induced by any dynamic persuasion rule must satisfy the L-principle (13). However, part (ii) shows that the L-principle (13) is *not sufficient* to characterize optimal persuasion rules in a large class of dynamic persuasion environments. This applies even within the restricted class of (k, n)-verification problems. The L-principle is only sufficient when $n - k = 1$.

4. Credibility

Glazer and Rubinstein (2006) show that in the static persuasion problem, there is *no value to commitment* for the listener: In the game where the speaker moves first, sending a message $m \in \sigma(x)$, and then the listener responds with an action, there is an equilibrium that leads to the same outcome as the optimal rule. This is the *credibility result*. Glazer and Rubinstein (2004) present a similar credibility result for dynamic persuasion rules for the special case of $(1, 2)$-verification environments described in the previous section. In this section, we extend the credibility result to general dynamic persuasion rules. Sher (2011) extends the static credibility result to nonbinary decisions under a concavity assumption.

In a dynamic persuasion problem, the *game without commitment* is a game that has the same timing as the dynamic communication protocol of Section 2.2. The difference is that the listener does not commit to a dynamic persuasion rule prior to the game, but rather decides which message to request (in step 3(a) of Figure 2), whether to reject without requesting further evidence (in step 3(b) of Figure 2), and whether to accept or reject conditional on the receipt of any message (in step 5 of Figure 2). In particular, in step 5 of Figure 2, the listener is not forced to accept the speaker's request if the speaker presents the message that the listener requested. Neither is the listener forced to reject the speaker's request if some other message is presented.

Define the speaker's *reporting strategy* to be the part of the speaker's (behavioral) strategy that specifies her cheap talk claim about the state. Let $\zeta(x, \hat{x})$ be the probability that the speaker claims that the state is $\hat{x}$ conditional on the actual state being x. To attain a complete description of the speaker's strategy, we must supplement the reporting strategy ζ with a description of what evidence the speaker would present at step 4 of Figure 2 conditional on the previous history.

To analyze the credibility of the optimal persuasion rule, it is useful to analyze the dual of Program P0.

Program D. Maximize

$$\sum_{x\in A} \varphi(s,x) \tag{14}$$

subject to

$$\varphi(s,x) \le p_x \quad \forall x \in A \tag{15}$$

$$\varphi(y,t) \le p_y \quad \forall y \in R \tag{16}$$

$$\varphi(s,x) \le \sum_{y\in R: m\in\sigma(y)} \varphi(x,y) \quad \forall x \in A, \forall m \in \sigma(x) \tag{17}$$

$$\varphi(y,t) = \sum_{x\in A} \varphi(x,y) \quad \forall y \in R \tag{18}$$

$$\varphi(x,y) \ge 0 \quad \forall x \in A, \forall y \in R. \tag{19}$$

For each $x \in A$ and $y \in R$, there is a variable $\varphi(x,y)$ that is the multiplier on the incentive constraint (5) corresponding to the pair (x,y), which says that at y, the speaker must achieve at least the acceptance probability that she would achieve if she were to claim that the state were x. In addition, for each $x \in A$, there is a variable $\varphi(s,x)$ that is the multiplier on the constraint (4), the constraint that equates the error probability at x with the probability that the speaker is rejected at x given that she reports truthfully. The reason for the notation s in $\varphi(s,x)$ is explained in Section 5 below. For each $y \in R$, there is a variable $\varphi(y,t)$, where again the notation t is explained in Section 5. The variable $\varphi(y,t)$ does not correspond to any constraint in Program P0, but is rather a variable defined by (18). The introduction of this variable simplifies the constraint (16). We will use the notation $\varphi = (\{\varphi(s,x)\}_{x\in A}, \{\varphi(x,y)\}_{(x,y)\in A\times R}, \{\varphi(y,t)\}_{y\in R})$ to denote a feasible solution in Program D.

Define a *credible implementation* of a dynamic persuasion rule g to be a sequential equilibrium of the game without commitment that induces error probability $\mu_x(g)$ at each state x. In other words, the credible implementation leads to the same outcome as committing to g.

Theorem 5. *Let (μ^*, β^*) be an optimal solution to Program P0, let g^* be the corresponding optimal dynamic persuasion rule defined by (8), and let φ^* be an optimal solution to Program D. Then there exists a credible implementation of g^*. The speaker's reporting strategy is*

$$\zeta^*(x,y) := \begin{cases} 1 & \text{if } x \in A \text{ and } x = y \\ \varphi^*(y,x)/\varphi^*(x,t) & \text{if } x \in R \text{ and } y \in A \\ 0 & \text{otherwise,} \end{cases} \tag{20}$$

except at $x \in R$ with $\varphi^*(x,t) = 0$, where $\zeta^*(x,y)$ can be chosen arbitrarily. On the equilibrium path, the speaker presents the evidence the listener requests whenever it is available, whereas the listener's behavior coincides with what it would have been if he had committed to g^* as in Section 2.2.

Theorem 5 is a credibility result for *dynamic* persuasion rules. The theorem shows that optimal dynamic persuasion rules derived from Program P0 are credible in the sense that they are consistent with equilibrium in the game without commitment. The credible implementation achieves the same error probability as the optimal rule in the game with commitment. The theorem also shows that there always exists an equilibrium that supports the optimal rule at which, in A, the speaker tells the truth, while in R, the speaker randomizes over lies.

Theorem 5 extends the result of Glazer and Rubinstein (2004), which applies only to $(1,2)$-verification environments, to the class of *all* persuasion environments. Theorem 5 also complements the credibility result for *static* persuasion rules in Glazer and Rubinstein (2006). As in Glazer and Rubinstein (2004), the proof involves optimality conditions for a pair of dual linear programs. In our case, these programs are Program P0, from which the optimal rule is derived, and its dual, Program D, from which the speaker's strategy is derived. In contrast, Glazer and Rubinstein (2004) employed Program GR0 and its dual. However, Theorem 4 shows that Program GR0 is generally valid only for (k,n)-verification problems with $n-k=1$.

5. Foresight and Max Flow

We present a condition—foresight—under which the optimal persuasion problem reduces to the classical maximum flow problem. One consequence is that the optimal static rule is optimal even among dynamic rules. We derive additional consequences, translating properties of max flow to properties of optimal persuasion. In addition to the consequences we derive (determinism, coincidence of static and dynamic optimality, nature of credible equilibrium strategies, symmetry, and complexity), the reduction implies comparative statics, which we explore elsewhere (Sher 2008).

5.1 *Foresight*

We motivate foresight with an example.

Example 2. Let $M=\{1,\dots,n\}$. Interpret a message $m\in M$ as evidence that would be persuasive if it is *genuine* and not *fabricated*. In states $x\in A$, all evidence is genuine as the speaker is unwilling or unable to fabricate. For $x\in A$, we allow $\sigma(x)$ to be an arbitrary nonempty subset of M. In states $y\in R$, all evidence is fabricated. In general, no observable characteristic distinguishes genuine from fabricated evidence. Assume that if $i<j$, then evidence i is easier to fabricate than evidence j, which means that

$$\forall y\in R\quad i<j\quad\Rightarrow\quad (j\in\sigma(y)\Rightarrow i\in\sigma(y)). \tag{21}$$

$\Diamond$

The example is clearly special, yet any persuasion environment may be interpreted along similar lines: Define evidence m to be (weakly) *less forgeable* than m' if

$$\forall y\in R,\quad m\in\sigma(y)\quad\Rightarrow\quad m'\in\sigma(y).$$

Evidence m is *minimally forgeable* at $x \in A$ if $m \in \sigma(x)$ and m is weakly less forgeable than any $m' \in \sigma(x)$.[9]

Definition 1. A persuasion environment *allows foresight* if at each accept state $x \in A$, there exists a minimally forgeable message m_x.

We use the term "foresight" because if there is a minimally forgeable message at accept state x, then the speaker should foresee that this is the evidence that the listener would require to support the claim that the state is x (see Corollaries 1 and 2 below). For this reason we also refer to m_x as the *obvious question* at x. In contrast, when foresight fails, the listener may randomize, making evidence requests unpredictable.

Example 2 allows foresight, with minimally forgeable message $m_x = \max\{i : i \in \sigma(x)\}$ at $x \in A$. Unlike the example, foresight does not generally require a global ranking of evidence in terms of forgeability; foresight is even consistent with the availability of two messages at accept state x, neither of which is weakly more forgeable than the other. The (k, n)-verification environments (Section 3.2) typically violate foresight (depending on the choice of A).

Foresight generalizes *normality*, a notion used by Bull and Watson (2007), which is equivalent to the *full reports condition* of Lipman and Seppi (1995) and resembles the *nested range condition* of Green and Laffont (1986).[10] To compare normality to foresight, we generalize the forgeability relation. In particular, for any $Y \subseteq X$, define the relation $\preccurlyeq_Y$ on the set of messages by

$$m \preccurlyeq_Y m' \quad \Leftrightarrow \quad [\forall y \in Y, m \in \sigma(y) \Rightarrow m' \in \sigma(y)].$$

Say that m is $\preccurlyeq_Y$-minimal at x if (i) $m \in \sigma(x)$ and (ii) $m \preccurlyeq_Y m'$ for all $m' \in \sigma(x)$. So σ allows foresight if there is a $\preccurlyeq_R$-minimal message at every state in A. In contrast, σ is *normal* if there is a $\preccurlyeq_X$-minimal message at every state in X. So foresight is weaker than normality in two ways. First, foresight employs the finer relation $\preccurlyeq_R$, while normality employs the coarser relation $\preccurlyeq_X$: any $\preccurlyeq_X$-minimal message is $\preccurlyeq_R$-minimal, but not every $\preccurlyeq_R$-minimal message is $\preccurlyeq_X$-minimal, so that $\preccurlyeq_X$-minimal messages may fail to exist in states with $\preccurlyeq_R$-minimal messages. Hence, the requirement of the existence of $\preccurlyeq_X$-minimal messages is more stringent. Second, foresight only requires $\preccurlyeq_R$-minimal messages at *accept* states (that is, states in A), while normality requires $\preccurlyeq_X$-minimal messages at *all* states in X.

The interpretation of the two notions is also different: If a message m is $\preccurlyeq_X$-minimal at x, it is *maximally informative* in the sense that message m rules out (weakly) more

[9] The minimally forgeable message, when it exists, may not be unique.

[10] Papers that explore similar notions include Deneckere and Severinov (2008), Singh and Wittman (2001), and Forges and Koessler (2005). Models with normal message structures include Milgrom (1981), Milgrom and Roberts (1986), Shin (1994a, 1994b, 2003), Bull and Watson (2004), and Dziuda (2011). Ben-Porath and Lipman (2012) and Kartik and Tercieux (2012) study full implementation with evidence and explore normality in that context. An older version of the latter discusses when dynamic mechanisms outperform static mechanisms for full Nash implementation with evidence as well as the bearing of normality on this issue.

states (in the sense of inclusion) than any other message at x. This concords with the common interpretation according to which normality holds in the absence of time constraints, where the maximal message corresponds to a presentation of all available evidence. In contrast, a message that is $\preccurlyeq_R$-minimal cannot be said to be maximally informative in the above sense, as $\preccurlyeq_R$-minimality of m at x implies nothing about the relative availability of m at states (other than x) in A. The $\preccurlyeq_R$-minimality of m at x means that x is maximally informative with respect to the specific question of what the state could be *if the state belongs to* R. From another point of view, a $\preccurlyeq_X$-minimal message is one that it is difficult to *use* in the sense that it is (relatively) infrequently available, whereas a $\preccurlyeq_R$-minimal message is a message that it is difficult to *abuse* in the sense that it is difficult to use when the listener would not like to be persuaded. This justifies the terminology "forge" above.

For $Y \subseteq X$, consider the environment where the set of states is Y and the message correspondence is the restriction of σ to Y. If this environment is normal, then we say that the message correspondence σ is *normal on* Y. The following remark and its proof show how to construct many message correspondences that allow foresight but are not normal.

Remark 1. Any normal environment must be normal on Y for all $Y \subseteq X$. In contrast, (i) any message correspondence that is not normal on A can be extended to a message correspondence on X that allows foresight and (ii) any message correspondence that is not normal on R can be extended to a message correspondence on X that allows foresight.

For (i), let σ be arbitrary on A, assume without loss of generality that $M = \{1, \dots, n\}$, and extend σ to R so that it satisfies (21). Statement (ii) is illustrated by the following example.

Example 3. Let σ be arbitrary on R, but assume that for all $x \in A$ that $\sigma(x)$ contains only a single message.[11] Such environments allow foresight. For interpretation, assume the listener would like to accept the speaker only if she is *honest*. The speaker is honest if she is an essentially nonstrategic player—a behavioral type—whose only option is to present truthful evidence. In contrast, the listener would like to reject the speaker if she is *strategic*, and selects her evidence to maximize her benefit. $\diamondsuit$

While foresight entails an obvious question at every accept state, it does not imply that the optimal rule is obvious. Optimality requires that we determine the level at which evidence is sufficiently difficult to fabricate: If the speaker's opening argument claims that the state is $x \in A$, is m_x sufficiently difficult to forge for it to be worthwhile for the listener to ask the obvious question rather than rejecting the speaker outright? Section 5.2 determines whether m_x meets this criterion by solving a max flow problem (see Corollary 2).

[11] Strictly speaking, we assume that $\sigma(x)$ contains only a single nonvacuous message (see Section 2.1).

State	A/R	Prob	Messages
1	A	$\frac{3}{16}$	m_1, m_2, m_3
2	A	$\frac{2}{16}$	m_2, m_3
3	R	$\frac{1}{16}$	m_3
4	A	$\frac{1}{16}$	m_1, m_3, m_4, m_5
5	A	$\frac{2}{16}$	m_3, m_5
6	R	$\frac{3}{16}$	m_2, m_3, m_6
7	R	$\frac{1}{16}$	m_1, m_2, m_3, m_6
8	R	$\frac{3}{16}$	m_1, m_2, m_3, m_5

TABLE 1. A persuasion environment that allows foresight.

5.2 *Max Flow*

Under foresight, we reduce optimal persuasion to the classical maximum flow problem.[12] We begin by constructing a network that corresponds to the persuasion problem.[13] A directed graph is a pair (V, E), where V is a set of vertices and E is a set of directed edges. Assume $V = X \cup \{s, t\}$, where X is the set of states in the persuasion problem, and s and t are two new vertices. The vertex s is called the *source* and t is called the *sink*. The edge set is given by

$$E = \underbrace{\{(s, x) : x \in A\}}_{E_1} \cup \underbrace{\{(x, y) : x \in A, y \in R, \sigma(x) \subseteq \sigma(y)\}}_{E_2} \cup \underbrace{\{(y, t) : y \in R\}}_{E_3}. \tag{22}$$

In addition, each edge $(v, w) \in E$ has a *capacity* $c(v, w)$, defined as

$$c(v, w) := \begin{cases} p_x & \text{if } (v, w) = (s, x) \in E_1 \\ \infty & \text{if } (v, w) = (x, y) \in E_2 \\ p_y & \text{if } (v, w) = (y, t) \in E_3. \end{cases} \tag{23}$$

EXAMPLE 4. Table 1 describes a persuasion environment. The first column represents the set of states $X = \{1, \ldots, 8\}$, the second column shows whether each state belongs to A or R, the third column gives the probability of each state, and the last column gives the messages available at the state. One can verify that the environment specified by Table 1 allows foresight. At each state $x \in A = \{1, 2, 4, 5\}$, the obvious question is m_x. Figure 3 presents a network that corresponds to this persuasion environment. States in A and R correspond to vertices in the left and right columns, respectively. The graph also contains a source s and a sink t. An edge (in E_1) with capacity p_x points from s to each $x \in A$; an edge (in E_3) with capacity p_y points from each $y \in R$ to t. For all $x \in A$ and $y \in R$, if y can mimic x (i.e., $\sigma(x) \subseteq \sigma(y)$), there is an infinite capacity edge (in E_2) that points from x to y. Infinite capacities are not labeled to avoid cluttering the diagram. ◊

[12] For an extensive treatment of the maximum flow problem, see Ahuja et al. (1993).

[13] This resembles the conversion of the maximal closure problem to the max flow problem in Picard (1976).

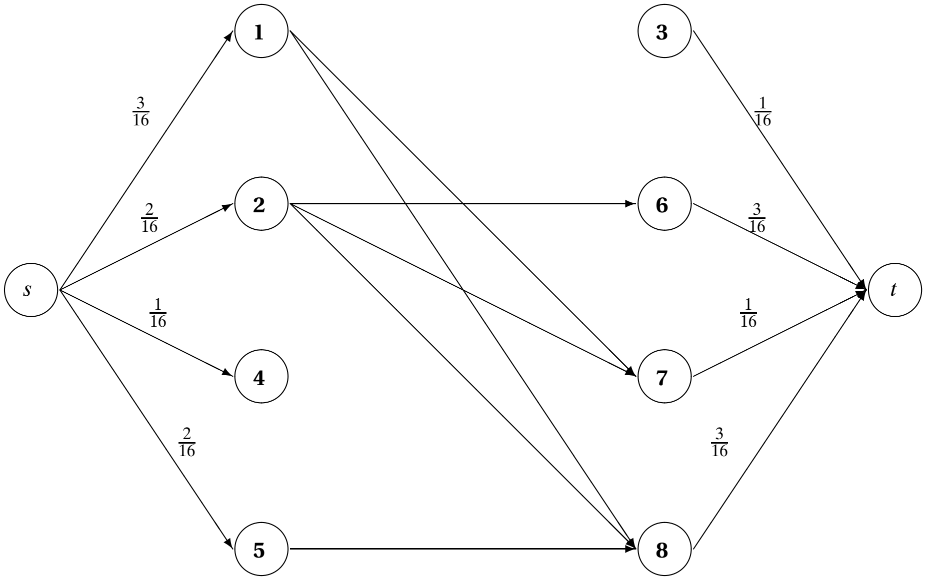

FIGURE 3. A network that corresponds to a persuasion environment.

The maximum flow problem on the above defined network is as follows.

MAX FLOW. Maximize

$$\sum_{x:(s,x)\in E} \varphi(s,x) \tag{24}$$

subject to

$$\sum_{v:(v,x)\in E} \varphi(v,x) = \sum_{v:(x,v)\in E} \varphi(x,v) \quad \forall x \in X\ (= V \setminus \{s,t\}) \tag{25}$$

$$\varphi(v,w) \le c(v,w) \quad \forall (v,w) \in E \tag{26}$$

$$\varphi(v,w) \ge 0 \quad \forall (v,w) \in E. \tag{27}$$

The variable $\varphi(v,w)$ may be conceived as a flow of some commodity along edge (v,w). The set of equations (25) comprises the *flow conservation constraints* into any vertex (other than the source or sink) is equal to the flow out of that vertex. Equation (26) comprises *capacity constraints* that say that the flow along any edge cannot exceed that edge's capacity. Max Flow attempts to ship as much flow out of the source as possible subject to these constraints. An optimal solution is a *maximum flow*. The formulation (24)–(27) is a special case of the maximum flow problem because the edge set has structure (22) and the capacity function takes the form (23).[14]

The dual of the maximum flow problem is the *minimum cut* problem.

[14]Typically, the max flow objective is the difference between the outflow and the inflow of the source; because here the edge set (22) does not contain edges that enter the source, the objective is simply the outflow.

MIN CUT. Minimize

$$\sum_{(v,w)\in E} \delta(v,w)c(v,w) \tag{28}$$

subject to

$$\delta(v,w) - \gamma_v + \gamma_w \geq 0 \quad \forall (v,w) \in E \tag{29}$$

$$\delta(v,w) \geq 0 \quad \forall (v,w) \in E \tag{30}$$

$$0 \leq \gamma_v \leq 1 \quad \forall v \in V \setminus \{s,t\} \tag{31}$$

$$\gamma_s = 1 \tag{32}$$

$$\gamma_t = 0. \tag{33}$$

The formulation of the objective function in (28) uses the convention that $0 \times \infty = 0$.[15] Strictly speaking, (31) is not part of the dual of (24)–(27), but imposing (31) does not affect the value of Min Cut. (See Lemma 1 in the Appendix.) An interpretation of Min Cut is provided below.

THEOREM 6. *Assume foresight. Then the following statements hold:*

(i) Any optimal solution to Max Flow is an optimal solution to Program D.[16]

(ii) Any optimal solution to Min Cut induces an optimal solution to Program P0 via

$$\begin{aligned} \beta(x,m) &= \begin{cases} \gamma_x & \text{if } x \in A \text{ and } m = m_x \\ 0 & \text{otherwise} \end{cases} \\ \mu_x &= \begin{cases} 1-\gamma_x & \text{if } x \in A \\ \gamma_x & \text{if } x \in R. \end{cases} \end{aligned} \tag{34}$$

Moreover, Max Flow, Min Cut, Program P0, and Program D all attain the same value.

Say that optimal dynamic and static persuasion rules *coincide* if optimal dynamic and static rules lead to the same error probability. It is well known that Min Cut always has an integer optimum solution, meaning one with $\gamma_x \in \{0,1\}$ for all $x \in V$ and $\delta(v,w) \in \{0,1\}$ for all $(v,w) \in E$.[17] Combining Theorems 1 and 2 with Theorem 6 immediately yields a corollary.

COROLLARY 1. *Assume foresight. Then there exists an optimal dynamic persuasion rule that is deterministic, and, hence, optimal dynamic and static rules coincide.*

While foresight implies that the optimal dynamic rule can be implemented using a static rule, it does *not* imply that *all* dynamic rules can be implemented using static

[15] We could have replaced ∞ by any sufficiently large real number in (23).

[16] Since Max Flow does not contain variables $\varphi(x,y)$ when $(x,y) \in (A \times R) \setminus E$, to interpret a Max Flow solution φ as a solution to Program P0, we extend φ to edges $(x,y) \in (A \times R) \setminus E$ by setting $\varphi(x,y) = 0$.

[17] This is a consequence of the total unimodularity of the constraint matrix.

rules. Specifically, modifying Example 1 so that either $A=\emptyset$ or $A=X$, then, vacuously, the example allows foresight. Corollary 1 then implies there is a static rule that is optimal even among dynamic rules. Indeed, if $A=\emptyset$, this is the static rule that rejects all messages, and if $A=X$, it is the static rule that accepts all messages. Defining $Z:=\{x:\sum_{i=1}^{3}x_i \text{ is even}\}$, the dynamic rule

$$g(x,m):=\begin{cases}\frac{1}{3} & \text{if } x\in Z \text{ and } m\in\sigma(x)\\ 0 & \text{otherwise}\end{cases} \tag{35}$$

induces a vector of acceptance probabilities $(\alpha(g,x):x\in X)$ that cannot be replicated by any static rule. Rule (35) is the same as (12) except that Z is no longer equal to A. Similarly, in Example 4, which allows foresight, consider the dynamic rule g with $g(x,m)=\frac{1}{2}$ if $x=1$ and $m\in\{m_1,m_2\}$, and $g(x,m)=0$ otherwise. No static rule mimics g (in the sense of inducing the same acceptance probabilities). Still, Corollary 1 entails the existence of a static rule in Example 4 that is optimal even among dynamic rules.

These examples establish that Corollary 1 is fundamentally different from Theorem 5 of Bull and Watson (2007), which shows that under normality, dynamic and static (weak) implementation coincide in their general mechanism design framework with evidence. Whereas their result presents a sufficient condition for anything feasible via a dynamic mechanism to be feasible via a static mechanism, our result establishes a sufficient condition for something optimal via a dynamic mechanism to be feasible, and hence optimal, via a static mechanism.[18]

We now interpret Min Cut. A *cut* is a set of vertices $Z\subseteq V$ with $s\in Z$, $t\notin Z$. Let $\mathcal{C}$ be the set of all cuts in our network. The *capacity of a cut* Z is

$$c(Z):=\sum\{c(v,w):v\in Z, w\in V\setminus Z,(v,w)\in E\}.$$

So the capacity of a cut Z is the sum of capacities of all edges crossing the cut. The set of cuts corresponds to the set of integer solutions to Min Cut. In particular, $\gamma_x=1$ exactly if $x\in Z$ and $\gamma_x=0$ otherwise; $\delta(v,w)=1$ if (v,w) crosses Z (i.e, $v\in Z$, $w\notin Z$) and $\delta(v,w)=0$ otherwise. Min Cut can then be rewritten as minimize $c(Z)$ subject to $Z\in\mathcal{C}$. An optimal solution Z is a *minimum cut*. Any maximum flow φ induces a minimum cut Z; the induced minimum cut is the set of vertices reachable from the source in the residual graph induced by φ (for details, see Ahuja et al. 1993).

Corollary 2. *Assume foresight. A maximum flow induces both an optimal rule and its credible implementation. Specifically, any minimum cut Z corresponds to the optimal rule*

$$g(x,m)=\begin{cases}1 & \textit{if } x\in A\cap Z \textit{ and } m=m_x\\ 0 & \textit{otherwise.}\end{cases} \tag{36}$$

[18]Another fundamental difference between our analysis and that of Bull and Watson (2007) is that whereas the full optimality of static rules among dynamic rules, as in Corollary 1, depends crucially on whether the listener can benefit by randomizing evidence requests, Bull and Watson (2007) do not discuss randomization at all as a determinant of whether static mechanisms are optimal.

The rule g accepts the speaker at states in $Z \setminus s$ and rejects the speaker otherwise. A maximum flow corresponds to a speaker strategy via (20).[19]

The optimal static rule that corresponds to the dynamic rule (36) is f such that $f(m) = 1$ if $m = m_x$ for $x \in A \cap Z$ and $f(m) = 0$ otherwise. Here, the credibility result can be rephrased in the context of the static communication protocol of Section 2.2.[20] The optimal static rule f gives the listener's strategy in the game without commitment,[21] and the listener's strategy is to send m_x when $x \in A$ and to randomize over messages $(m_y : y \in A)$ with probabilities given by (20) when the state x is in R. The Appendix applies Corollary 2 to solve Example 4, deriving the optimal rule and its credible implementation from the maximum flow.

Theorem 6 has consequences for the structure of the set of *optimal static persuasion rules*, meaning the set of static persuasion rules that are optimal *among* static rules. The following example helps us to explain this.

Example 5. This example is due to Glazer and Rubinstein (2006). Consider a $(2,5)$-verification environment with $X_i = \{0,1\}$ for $i = 1, \dots, 5$. (See Section 3.2 for the definition of (k,n)-verification environments.) Let $A := \{x \in \{0,1\}^5 : \sum_{i=1}^5 x_i \geq 3\}$ and $p_x = \frac{1}{32}$ for all x. Interpret $x_i = 1$ (resp., $x_i = 0$) as a fact that supports (resp., opposes) the speaker's request. So the listener prefers to accept exactly when a majority of facts support the speaker, but the speaker can present only two facts. The optimal static rule partitions the indices $\{1,2,3,4,5\}$ into two sets of size 2 and 3, which we call *categories*—for example $\{1,2\}$ and $\{3,4,5\}$—and accepts the speaker only if the speaker presents two facts that support her request *in the same category*. For example, if the evidence consists of the opinions of five experts, two of whom are women and three of whom are men, it would be optimal to require the speaker to present supporting opinions of two experts *of the same gender* to win acceptance. Categories such as gender are ex ante irrelevant, but the optimal rule may have to make use of some such categories. The model of Fishman and Hagerty (1990) on optimal disclosure has a similar character. $\Diamond$

We now formalize the asymmetry of the optimal rule found in the above example.

Definition 2. A pair of bijections (π, ξ) with $\pi : X \to X$ and $\xi : M \to M$ is a *symmetry* if for all x, (i) $\sigma(\pi(x)) = \{\xi(m) : m \in \sigma(x)\}$, (ii) $x \in A \Leftrightarrow \pi(x) \in A$, and (iii) $p_{\pi(x)} = p_x$. Static rule f is *symmetric* if for every symmetry (π, ξ) and every $m \in M$, $f(m) = f(\xi(m))$.

Intuitively, a persuasion rule is symmetric if it treats any two pieces of evidence that cannot be distinguished without labels—or, in other words, in terms of their intrinsic

[19] See Theorem 5 and its proof for the exact details of this strategy. See also footnote 16.

[20] Glazer and Rubinstein (2006) establish a credibility result for static rules. Our result differs in two ways: (i) the optimal static rule is also an optimal dynamic rule, so one can actually implement the optimal dynamic rule in the static communication protocol without commitment, and (ii) one can derive the credible implementation by solving a maximum flow problem. Neither (i) nor (ii) hold without foresight.

[21] We modify f to accept additional messages available only in A (if any exist) to ensure off-path rationality.

properties—in the same way. One can show that *every* optimal static rule in Example 5 is asymmetric. For this, it is critical that Example 5 violates foresight.

Theorem 6 allows us to identify a symmetric optimal static rule under foresight. Say a static rule f is *less difficult* than static rule f' if $f'(m) \leq f(m)$ for all $m \in M$. The less difficult is a rule f, the easier it is for the speaker to persuade the listener under f. A *least difficult optimal static persuasion rule* is a persuasion rule that is both optimal among static rules and less difficult than all other optimal static rules. Theorem 6 implies the following corollary.

Corollary 3. *Assume foresight. Then there exists a least difficult optimal static rule f^*. The rule f^* is symmetric.*

This result follows from the well known lattice structure of the set of minimum cuts (Ford and Fulkerson 1956). Under foresight, optimal deterministic static persuasion rules—or, more precisely, the acceptance probabilities induced by such rules—correspond to minimum cuts, where rules higher up on the lattice tend to make persuasion easier. The rule f^* is the persuasion rule that accepts message m exactly if m is accepted by some optimal deterministic static rule. Clearly f^* is symmetric, but as shown by Example 5, in general, f^* need not be optimal. However, $(\alpha(f^*, x) : x \in X)$ is the supremum of the acceptance probabilities of the set of optima, implying *under foresight* that since the set of optima is a lattice, f^* is itself optimal.

In contrast to static rules, there always (even without foresight) exists a symmetric optimal dynamic rule g^* due to the convexity of the set of acceptance probability vectors induced by dynamic rules. However, unlike in the case of foresight, where f^* is deterministic, without foresight, there need not exist a deterministic symmetric optimal dynamic rule. In light of Example 5, this suggests that in the absence of foresight, *some arbitrariness in categorization can be a (imperfect) substitute for randomized back-and-forth communication*. Also, g^* need not be the least difficult optimal dynamic rule as no such rule may exist. However, under foresight, f^* is optimal even among dynamic rules, and at every state, the speaker (weakly) prefers f^* to all dynamic rules.[22]

6. Complexity

Theorems 1 and 6 have consequences for the complexity of optimal persuasion.[23]

Theorem 7. *(i) The static persuasion problem is NP-hard.*

(ii) In contrast, for environments that allow foresight, there is a polynomial time algorithm that computes both an optimal static rule f and a credible implementation of f.

(iii) Even without foresight, there exists a polynomial time algorithm that computes both an optimal dynamic persuasion rule g and a credible implementation of g.

[22] I am grateful to an anonymous referee for emphasizing this point.

[23] See, for example, Goldreich (2008) for definitions of the standard complexity notions employed here.

Part (iii) follows from the fact that the dynamic persuasion problem can be represented by the linear Program P0 (Theorem 1) and the speaker strategy in credible implementation of the optimal rule can be found by solving the dual linear Program D (Theorem 5).[24] Part (ii) follows similarly from Theorem 6 on reduction of optimal persuasion to the linear program Max Flow under foresight, Corollaries 1 and 2, and the discussion following Corollary 2 about credible implementation in the static protocol. Note that whether a persuasion environment allows foresight can also be determined in polynomial time. The proof of part (i) is given in the Appendix and is by reduction from the set cover problem.

The result that with either (i) dynamic communication or (ii) foresight, not only is finding the optimal rule g tractable, but so is finding an equilibrium supporting g, contrasts with the general result that finding a Nash equilibrium in an arbitrary game is hard (formally, PPAD-complete) (Daskalakis et al. 2009, Chen and Deng 2006) and that finding the optimal Nash equilibrium for a given player is NP-hard (Gilboa and Zemel 1989). That the listener optimal equilibrium is derived via a pair of dual linear programs resembles the fact that in zero-sum games, an equilibrium can be derived via a pair of dual linear programs,[25] and that the optimal *correlated equilibrium* in a normal form game for any given player can be derived via a linear program (Gilboa and Zemel 1989).[26]

7. Conclusion

This paper provides a new formulation of the optimal persuasion problem (Theorems 1 and 2). We thereby characterize the precise relation between static and dynamic persuasion. The methodology of this paper is promising for various extensions of the persuasion scenario and also for a broader class of problems concerning mechanism design with evidence.[27] Sher and Vohra (2013) apply a similar methodology to the problem of optimal price discrimination on the basis of evidence, a problem that has strong ties to optimal auctions. That paper shows that the optimal price discrimination problem reduces to a minimum convex cost flow problem. Analogously to the current paper, the reduction to a network flow problem in Sher and Vohra (2013) is used to derive a credibility result for the price discrimination environment. That paper also establishes a surprisingly close relationship between the price discrimination and persuasion problems.

[24]Both Programs P0 and D are easy to construct given the primitives of the problem, and they have a number of variables and constraints that is polynomial in the number of states and messages. The proof of Theorem 5 implies that the parts of the speaker strategy other than the reporting strategy are also easy to construct.

[25]In zero-sum games, all equilibria yield the same payoffs to all players, so finding any equilibrium is equivalent to finding the best equilibrium for any given player. Sher (2013) explores the relation between persuasion games and zero-sum games in more depth.

[26]For more results on the complexity of equilibria, see Conitzer and Sandholm (2008).

[27]For a general treatment of the linear programming approach to mechanism design and, in particular, the application of network flow problems to mechanism design, see Vohra (2011).

The specific concept of foresight also promises to generalize to a broader class of mechanism design problems. While the definition of normality treats all states symmetrically (see Section 5.1), and so depends only on properties of the evidence, foresight treats evidence at accept and reject states differently. A notion that treats states symmetrically—such as normality—must impose a more stringent condition on evidence to guarantee that optimal static and dynamic mechanisms coincide. But foresight appears to be very closely tailored to the persuasion environment. How might it be generalized for problems other than persuasion? The key to understanding why normality can be weakened to foresight in the persuasion environment is the observation that the only binding incentive constraints are those that say that speaker types in reject states should not want to mimic speaker types in accept states. This accounts for the role in foresight of forgery of evidence (i.e., use of evidence at a reject state that is available at an accept state), as opposed to, for example, the use of evidence available at one accept state at another accept state. Suppose that in a more general mechanism design problem with evidence, one can argue a priori, without knowing the precise structure of evidence, that at the optimal mechanism, all binding incentive constraints that involve mimicking any type t concern the mimicking of t by some type s in a subset $S(t)$ of the set of types.[28] (In an auction setting, $S(t)$ might be the set of types with higher values than t.) Then one could generalize foresight:[29] A message is t-forged whenever it is used by a type in $S(t)$, and the environment satisfies foresight if every type t has a message that is maximally difficult to t-forge (among messages available to t). Such a generalized notion of foresight is a promising avenue of exploration for more general mechanism design problems with evidence.

Appendix

Proof of Proposition 1. Fix mechanism (Γ, λ) with $\Gamma \in \mathbb{G}^n$ and fix a best reply to the mechanism for the speaker. For any $x \in X$ and $S \in \sigma^n(x)$, let $H(x, S)$ be the set of terminal histories h in $\Gamma(x)$ such that (i) the speaker is accepted at h and (ii) S is the set of messages m such that the speaker presents m at some evidentiary subhistory of h. Set $g(x, S)$ equal to the probability that at state x, the history belongs to $H(x, S)$ given that the speaker uses her best reply; this probability is calculated *conditional on* the state being x. It is a best reply to dynamic rule g for the speaker to report the state truthfully. If at some state x, the speaker had a strict incentive to lie to g and claim that the state was y, then she would have had a strict incentive to mimic y in (Γ, λ), contrary to the assumption that the speaker used a best reply to (Γ, λ). Indeed, mimicking y in (Γ, λ) can only be more attractive than claiming the state is y in g because in g, if the speaker fails to present the requested evidence $S \in \sigma^n(y)$, she will be rejected, whereas in (Γ, λ), if the speaker fails to present the evidence that would have led to acceptance at y, it is possible that she will be accepted. Given that the speaker reports truthfully to g, it is immediate that g and (Γ, λ) lead to the same total error probability. □

[28] A type here determines both the agent's evidence and any other relevant characteristics, such as, for example, the agent's value for an object in an auction.

[29] For this generalization to be nontrivial, it must be the case that at least for some types t, $S(t)$ is a proper subset of the set of types other than t.

Proof of Theorem 1. Consider an arbitrary dynamic rule g. For every state $x \in A$, let $h(x)$ be an optimal cheap talk report for the speaker against g at x. For all $x \in A$ and $m \in \sigma(x)$, define $\beta(x,m) := g(h(x),m)$ if $m \in \sigma(h(x))$ and $\beta(x,m) := 0$ otherwise. For $z \in X$, define $\mu_z := \mu_z(g)$. Then $\forall x \in A$, $\forall y \in R$, $\sum_{m\in\sigma(x)\cap\sigma(y)} \beta(x,m) = \sum_{m\in\sigma(x)\cap\sigma(h(x))\cap\sigma(y)} g(h(x),m) \le \sum_{m\in\sigma(h(x))\cap\sigma(y)} g(h(x),m) \le \mu_y(g) = \mu_y$. Using the definition of $h(x)$, it is now straightforward to verify that (β,μ) is feasible in Program P0. The equality $\mu_z := \mu_z(g)$ and the fact that g was arbitrary now imply that the value of Program P0 is a lower bound on the error probability of the optimal dynamic rule. So to complete the proof, it is sufficient to show that the rule g^* defined in (8), using an optimal solution (β^*,μ^*) to Program P0, satisfies $\mu_z(g^*) = \mu_z^*$ for all $z \in X$.

Optimality and (8) imply $\forall y \in R$, $\mu_y^* = \max\{\sum_{m\in\sigma(x)\cap\sigma(y)} \beta^*(x,m) : x \in A\} = \mu_y(g^*)$. Assume for contradiction that there exist $x', x'' \in A$ with

$$\sum_{m\in\sigma(x'')\cap\sigma(x')} \beta^*(x'',m) > \sum_{m\in\sigma(x')} \beta^*(x',m). \tag{37}$$

Define $\bar\beta(x,m) := \beta^*(x,m)$ except when $x = x'$. If $m \in \sigma(x'') \cap \sigma(x')$, define $\bar\beta(x',m) := \beta^*(x'',m)$. If $m \in \sigma(x') \setminus \sigma(x'')$, define $\bar\beta(x,m) := 0$. Define $\bar\mu_z := \mu_z^*$ except when $z = x'$. Define $\bar\mu_{x'} := 1 - \sum_{m\in\sigma(x')} \bar\beta(x',m)$. Using feasibility of (μ^*,β^*), it now follows that $\forall y \in R$, $\sum_{m\in\sigma(x')\cap\sigma(y)} \bar\beta(x',m) = \sum_{m\in\sigma(x')\cap\sigma(x'')\cap\sigma(y)} \bar\beta(x',m) \le \sum_{m\in\sigma(x'')\cap\sigma(y)} \beta^*(x'',m) \le \mu_y^* = \bar\mu_y$. With this inequality, it is straightforward to verify feasibility of $(\bar\mu,\bar\beta)$ in Program P0. Inequality (37) implies that $\bar\mu_{x'} < \mu_{x'}^*$, which in turn implies that $(\bar\mu,\bar\beta)$ attains a lower value than (μ^*,β^*) in Program P0, a contradiction. So $\forall x', x'' \in A$, $\sum_{m\in\sigma(x')} \beta^*(x',m) \ge \sum_{m\in\sigma(x'')\cap\sigma(x')} \beta^*(x'',m)$, implying that $\forall x' \in A$, $\mu_{x'}^* = 1 - \sum_{m\in\sigma(x')} \beta^*(x',m) = 1 - \sum_{m\in\sigma(x')\cap\sigma(x')} \beta^*(x',m) = 1 - \max\{\sum_{m\in\sigma(x)\cap\sigma(x')} \beta^*(x,m) : x \in A\} = \mu_{x'}(g^*)$. We have established that for all $z \in X$, $\mu_z(g^*) = \mu_z^*$, completing the proof. $\square$

Proof of Theorem 2. Proposition 1 of Glazer and Rubinstein (2006) says there exists an optimal solution to the static persuasion problem that is deterministic. So let f be an arbitrary deterministic static rule. For each $x \in A$, choose $m(x) \in \sigma(x)$ with $f(m(x)) = \alpha(f,x)$. For $x \in A$, define $\beta(x,m(x)) := f(m(x))$ and $\beta(x,m) := 0$ for $m \ne m(x)$. Define $\mu_z := \mu_z(f)$ $\forall z \in X$. Let $x \in A$ and $y \in R$. If $m(x) \in \sigma(y)$, then $\mu_y = \mu_y(f) \ge f(m(x)) = \sum_{m\in\sigma(x)\cap\sigma(y)} \beta(x,m)$. If $m(x) \notin \sigma(y)$, then $\mu_y = \mu_y(f) \ge 0 = \sum_{m\in\sigma(x)\cap\sigma(y)} \beta(x,m)$. With these inequalities, it is straightforward to verify that (μ,β) is feasible in Program P1. The equality $\mu_z := \mu_z(f)$ and the fact that f was arbitrary now imply that the value of Program P0 is a lower bound on the value of the static persuasion problem.

Let (μ^*,β^*) be optimal in Program P0. Constraints (4), (6), and (9) imply

$$\forall x \in A, \quad \sum_{m\in\sigma(x)} \beta^*(x,m) \in \{0,1\}. \tag{38}$$

Define static rule f^* by (11). For $x \in A$, $\mu_x(f^*) = 1 - \max\{f^*(m) : m \in \sigma(x)\} = 1 - \max\{\max\{\beta^*(x',m) : x' \in A\} : m \in \sigma(x)\} \le 1 - \max\{\beta^*(x,m) : m \in \sigma(x)\} = 1 - \sum_{m\in\sigma(x)} \beta^*(x,m) = \mu_x^*$, where the second to last equality follows from (9) and (38), and the last equality follows from the feasibility of (μ^*,β^*). Similarly, for $y \in R$,

$\mu_y(f^*) = \max\{f^*(m) : m \in \sigma(y)\} = \max(\{0\} \cup \{\beta^*(x,m) : x \in A, m \in \sigma(x) \cap \sigma(y)\}) = \max\{\sum_{m\in\sigma(x)\cap\sigma(y)} \beta^*(x,m) : x \in A\} \leq \mu_y^*$. The above implies part (i) of Theorem 2.

Let (μ^*, β^*) be optimal in Program P2. For each $x \in A$ with $\mu_x^* = 0$, (4) implies that there exists $m(x) \in \sigma(x)$ with $\beta^*(x, m(x)) > 0$. For $x \in A$, define $\beta^{**}(x,m) := 1$ if both $\mu_x^* = 0$ and $m = m(x)$; define $\beta^{**}(x,m) := 0$ otherwise.

Let $y \in R$. If $\mu_y^* = 0$, (5) and (7) imply that for all $x \in A$ and $m \in \sigma(x) \cap \sigma(y)$, $\beta^*(x,m) = 0$. So the definition of $m(x)$ implies $\forall x \in A$, $\mu_x^* = 1 \Rightarrow m(x) \notin \sigma(y)$. So $\forall x \in A$, $\mu_y^* = 0 = \sum_{m\in\sigma(x)\cap\sigma(y)} \beta^{**}(x,m)$. If $\mu_y^* = 1$, then $\forall x \in A$, $\mu_y^* = 1 \geq \sum_{m\in\sigma(x)\cap\sigma(y)} \beta^{**}(x,m)$. Using these relations, it is straightforward to verify that (μ^*, β^{**}) is feasible in Program P1.

Any optimal solution (μ, β) to Program P1 must satisfy

$$\mu_y = \max\left\{\sum_{m\in\sigma(x)\cap\sigma(y)} \beta(x,m) : x \in A\right\}.$$

This together with the Program P1 constraints implies (10) and, hence, that (μ, β) is feasible in Program P2. The above arguments imply that Programs P1 and P2 have the same value, and, moreover, imply part (ii) of Theorem 2. □

Proof of Theorem 4.

Proof of part (i) of Theorem 4. Let (x, T) be an L. Any feasible solution (μ, β) in Program P0 satisfies $1 - \mu_x = \sum_{m\in\sigma(x)} \beta(x,m) \leq \sum_{y\in T} \sum_{m\in\sigma(x)\cap\sigma(y)} \beta(x,m) \leq \sum_{y\in T} \mu_y$, where the first inequality follows from the fact that because (x, T) is an L, $\sigma(x) \subseteq \bigcup_{y\in T} \sigma(y)$. So the Program P0 constraints imply the L-principle and the result now follows from Theorem 1. ◁

Proof that (ii)(a) ⇒ (ii)(b). For $i = 1, \dots, n$, let $X_i = \{0, 1\}$. For all $z \in X = \times_{i=1}^n X_i$, let $p_z = 1/2^n$. Let $A = \{x : \sum_{i=1}^n x_i \text{ is even}\}$ (0 is even). Let $\mathcal{J} := \{J \subseteq \{1, \dots, n\} : |J| = k\}$. Then Program P0 simplifies to

$$\begin{aligned}
&\text{minimize} \\
&\frac{1}{2^n} \sum_{z\in X} \mu_z \\
&\text{subject to} \\
&\quad 1 - \mu_x = \sum_{J\in\mathcal{J}} \beta(x, x_J) \quad \forall x \in A \\
&\quad \mu_y \geq \sum_{J\in\mathcal{J}: y_J = x_J} \beta(x, x_J) \quad \forall x \in A, \forall y \in R \\
&\quad \mu_z \geq 0, \qquad \beta(x, x_J) \geq 0 \quad \forall z \in X, \forall x \in A, \forall J \in \mathcal{J}.
\end{aligned}$$

To understand this, note that if $|J| < k$ but $\beta(x, x_J) > 0$, then for some J' with $|J'| = k$ and $J' \supset J$, we can redefine $\beta(x, x_{J'}) := \beta(x, x_{J'}) + \beta(x, x_J)$ and $\beta(x, x_J) := 0$, maintaining feasibility as well as the objective function value.

Symmetry of the problem and convexity of the feasible region yields the existence of a symmetric optimal solution, i.e., an optimal solution (μ^*, β^*) such that there exist numbers μ_A^*, μ_R^*, and β^* with $\mu_x^* = \mu_A^*$ $\forall x \in A$, $\mu_y^* = \mu_R^*$ $\forall y \in R$, and $\beta^*(x, x_J) = \beta^*$ $\forall x \in A$, $\forall J \in \mathcal{J}$. So Program P0 further simplifies to

$$\begin{aligned}
&\text{minimize}\\
&\frac{1}{2}\mu_A + \frac{1}{2}\mu_R\\
&\text{subject to}\\
&\quad 1 - \mu_A = \binom{n}{k}\beta\\
&\quad \mu_R \geq \binom{n-1}{k}\beta \qquad\qquad (39)\\
&\quad \mu_A \geq 0, \qquad \mu_R \geq 0, \qquad \beta \geq 0.
\end{aligned}$$

To understand this, note that the tightest lower bound on μ_y for $y \in R$ is obtained using $x \in A$ that differs from y on exactly one component, in which case

$$|\{J \in \mathcal{J} : y_J = x_J\}| = \binom{n-1}{k}.$$

At optimality, (39) holds with equality, implying that

$$\mu_R = \left[\frac{\binom{n-1}{k}}{\binom{n}{k}}\right](1 - \mu_A) = \frac{n-k}{n}(1 - \mu_A),$$

so the objective value as a function of μ_A is

$$12\left(\mu_A + \frac{n-k}{n}(1 - \mu_A)\right) = \frac{n-k}{2n} + \frac{k}{2n}\mu_A,$$

which is minimized at $\mu_A = 0$. So the value of Program P0 is $(n-k)/2n$.

Choose $T \subseteq R$ with $|T| \leq k$ and $x \in A$. For any $y \in R$, there exists $i(y) \in \{1, \dots, n\}$ with $y_{i(y)} \neq x_{i(y)}$. Let $J = \{i(y) : y \in T\}$. Then $|J| \leq k$. So $x_J \in \sigma(x) \setminus \bigcup_{y \in T} \sigma(y)$, implying that (x, T) is not an L. It follows that

$$\text{if } (x, T) \text{ is a minimal } L, \quad \text{then } |T| \geq k + 1. \tag{40}$$

So define $\mu'_z := 1/(k+1)$ if $z \in R$ and $\mu'_z := 0$ if $z \in A$. The statement (40) implies that $\mu' = (\mu'_z : z \in X)$ satisfies all L constraints and, hence, is feasible in Program GR0. Moreover, μ' leads to an error probability of $1/[2(k+1)]$, which is less than $(n-k)/2n$—the value of Program P0—whenever $n - k > 1$ (and $0 < k$).[30] We have constructed a (k, n)-verification problem in which the value of Program GR0 is strictly less than the value of Program P0. Theorem 1 completes the proof. ◁

[30]Observe that when $k > 0$, $n > k + 1 \Rightarrow n - kn < n - k(k+1) \Rightarrow n < n(k+1) - k(k+1) \Rightarrow 1/(k+1) < (n-k)/n$.

Proof that (ii)(b) $\Rightarrow$ (ii)(a). Consider a (k,n)-verification problem with $n-k=1$. For any $z=(z_1,z_2,\ldots,z_n)\in X$, $i\in\{1,\ldots,n\}$, and $y_i\in X_i$, let $(y_i,z_{-i}):=(z_1,\ldots,z_{i-1},y_i,z_{i+1},\ldots,z_n)$. Let $\mu:=(\mu_z:z\in X)$ be feasible in Program GR0. Let $A'':=\{x\in A:\forall i,\exists y_i\in X_i,(y_i,x_{-i})\in R\}$. For all $x\in A''$ and $i\in\{1,\ldots,n\}$, choose $y_i^x\in\arg\min_{y_i\in X_i}\{\mu_{(y_i,x_{-i})}:(y_i,x_{-i})\in R\}$. The equality $n-k=1$ implies that $(x,\{(y_i^x,x_{-i}):i=1,\ldots,n\})$ is a minimal L, so that

$$\forall x\in A'',\quad \mu_x+\sum_{i=1}^{n}\mu_{(y_i^x,x_{-i})}\geq 1. \tag{41}$$

For all $x\in A':=A\setminus A''$, there exists $i(x)\in\{1,\ldots,n\}$ with $(y_{i(x)},x_{-i(x)})\in A\ \forall y_{i(x)}\in X_{i(x)}$. Let $J(i):=\{1,\ldots,n\}\setminus i$, so that $x_{J(i)}$ is the message that shows the value and index of all components other than i at x. Define

$$\beta'(x,m):=\begin{cases}1 & \text{if } x\in A' \text{ and } m=x_{J(i(x))}\\ \mu_{(y_i^x,x_{-i})}/\sum_{j=1}^n\mu_{(y_j^x,x_{-j})} & \text{if } x\in A'', m=x_{J(i)}, \text{ and } \sum_{j=1}^n\mu_{(y_j^x,x_{-j})}>1\\ \mu_{(y_i^x,x_{-i})} & \text{if } x\in A'', m=x_{J(i)}, \text{ and } \sum_{j=1}^n\mu_{(y_j^x,x_{-j})}\leq 1\\ 0 & \text{otherwise.}\end{cases}$$

Define $\mu'_z:=1-\sum_{m\in\sigma(x)}\beta'(x,m)$ if $z\in A$ and $\mu'_z:=\mu_z$ if $z\in R$. Let $x\in A$ and $y\in R$. There exists $i^*\in\{1,\ldots,n\}$ with $x_{i^*}\neq y_{i^*}$. If $y_{-i^*}=x_{-i^*}$, then $\mu'_y=\mu_y\geq\beta'(x,x_{J(i^*)})=\sum_{m\in\sigma(x)\cap\sigma(y)}\beta'(x,m)$. If $y_{-i^*}\neq x_{-i^*}$, then $\mu'_y\geq 0=\sum_{m\in\sigma(x)\cap\sigma(y)}\beta'(x,m)$. Using the above, it is straightforward to verify that (μ',β') is feasible in Program P0.

Choose $x\in A$. If either $x\in A'$ or $\sum_{j=1}^n\mu_{(y_j^x,x_{-j})}>1$, then $\mu'_x=0\leq\mu_x$. If $x\in A''$ and $\sum_{j=1}^n\mu_{(y_j^x,x_{-j})}\leq 1$, then $\mu'_x=1-\sum_{m\in\sigma(x)}\beta(x,m)=1-\sum_{i=1}^n\mu_{(y_i^x,x_{-i})}\leq\mu_x$, where the last inequality follows from (41). For $y\in R$, $\mu'_y=\mu_y$. So $\sum_{z\in X}p_z\mu'_z\leq\sum_{z\in X}p_z\mu_z$. Since μ was an arbitrary feasible solution in Program GR0, it follows that the value of Program P0 is a lower bound on the value of Program GR0. However, part (i) of the theorem now implies that the values of Programs P0 and GR0 are indeed the same. $\triangleleft$

$\square$

Proof of Theorem 5. First we specify speaker and listener strategies, and argue that they constitute a Bayesian Nash equilibrium (i.e., they are mutual best replies) that induces the error probability of the optimal rule. Afterward, we explain how these strategies can be modified to form a sequential equilibrium.

A Bayesian Nash equilibrium implements the optimal rule. The speaker uses the reporting strategy ζ^* given by (20) (unless $y\in R$ and $\varphi^*(y,t)=0$, in which case the reporting strategy selects an arbitrary probability distribution over reports in A[31]). At step 4 of Figure 2, the speaker presents the message the speaker requested if it belongs to $\sigma(x)\cap\sigma(\hat{x})$, where x is the true state and $\hat{x}$ was the speaker's cheap talk claim, and presents m^0 otherwise. The listener's strategy coincides with the optimal rule g^* given by (8). In particular, this means that at step 5 of Figure 2, the listener accepts the speaker exactly if the speaker presented the message that the listener requested (even though

[31] Nothing of substance would change if we allowed the speaker to also select reports in R in this case.

the rules of the game do not require the listener to do this). Call any listener strategy that follows this rule for behavior at step 5 of Figure 2 and also only requests messages $m \in \sigma(\hat{x})$ with positive probability following cheap talk report $\hat{x}$ *straightforward*. Call the speaker's proposed equilibrium strategy $\bar{\zeta}^*$ and the listener's strategy $\bar{g}^*$. Strategy $\bar{g}^*$ is straightforward.

Clearly, given the listener's strategy, the speaker would never have an incentive to modify her behavior at step 4 of Figure 2,[32] so we hold this behavior fixed throughout and consider only deviations from the speaker's reporting strategy.

Consider an arbitrary listener strategy $\bar{h}$. Let $h(\hat{x}, m)$ be the probability that the listener requests message m conditional on cheap talk report $\hat{x}$. (The listener rejects without requesting evidence with probability $1 - \sum_{m \in M} h(\hat{x}, m)$.) Let $h^\circ(\hat{x}, m, m')$ be the probability that the listener accepts the speaker conditional on the event that the speaker claims that the state is $\hat{x}$, the listener requests message m, and then the speaker presents message m'. Strategy $\bar{h}$ may not be straightforward. Let $\bar{g}$ be an alternative straightforward listener strategy in which, conditional on cheap talk report $\hat{x}$, the listener requests message m with probability

$$g(\hat{x}, m) := \begin{cases} h(\hat{x}, m) h^\circ(\hat{x}, m, m) \\ \quad + (1/|\sigma(\hat{x})|) \sum_{m' \in M \setminus \sigma(\hat{x})} h(\hat{x}, m') h^\circ(\hat{x}, m', m^0) & \text{if } m \in \sigma(\hat{x}) \\ 0 & \text{otherwise.} \end{cases}$$

Then $\bar{g}$ performs at least as well as $\bar{h}$ against $\bar{\zeta}^*$.[33] So in considering listener deviations from $\bar{g}^*$, we can restrict attention to straightforward strategies and, hence, consider deviations only from the listener's strategy for requesting evidence at step 3 of Figure 2.

The proof of Theorem 1 establishes that for any optimal solution (μ^*, β^*) to Program P0 and any $x, x' \in A$, $\sum_{m \in \sigma(x')} \beta^*(x', m) \geq \sum_{m \in \sigma(x'') \cap \sigma(x')} \beta^*(x'', m)$, implying that it is a best reply for the speaker to report truthfully at states $x \in A$.[34]

Let $y \in R$. Assume for contradiction that there exist $x, x' \in A$ with $\zeta^*(y, x) > 0$ but $\sum_{m \in \sigma(x') \cap \sigma(y)} g^*(x', m) > \sum_{m \in \sigma(x) \cap \sigma(y)} g^*(x, m)$. Then by (5), $\mu_y^* > 0$. Then complementary slackness implies $\varphi^*(y, t) = p_y > 0$. So at y, ζ^* is given by (20), implying $\varphi^*(x, y) > 0$. Complementary slackness now implies $\mu_y^* = \sum_{m \in \sigma(x) \cap \sigma(y)} g^*(x, m)$, implying that $\mu_y^* < \sum_{m \in \sigma(x') \cap \sigma(y)} g^*(x', m)$ and contradicting (5). So $\bar{\zeta}^*$ is a best reply to $\bar{g}^*$.

For $x \in A$ and $m \in \sigma(x)$, define $\widehat{p}_x := p_x + \sum_{y \in R} \zeta^*(y, x) p_y$ and define $\delta_{x,m} := \sum_{y \in R : m \in \sigma(y)} \zeta^*(y, x) p_y$. The quantity $\widehat{p}_x$ is the probability that the speaker will *claim* that the state is x according to ζ^*. The quantity $\delta_{x,m}$ is the probability that (i) the state is in R, (ii) the speaker claims that the state is x, and (iii) the speaker has message m.

[32] Recall that $g^*(\hat{x}, m) > 0 \Rightarrow m \in \sigma(\hat{x})$.

[33] Assume the speaker uses $\bar{\zeta}^*$. If $x \in A$, then regardless of whether the listener uses $\bar{h}$ or $\bar{g}$, the acceptance probability conditional on the state being x is $\sum_{m \in \sigma(x)} h(x, m) h^\circ(x, m, m) + \sum_{m' \in M \setminus \sigma(x)} h(x, m') h^\circ(x, m', m^0)$. Let $y \in R$. If the listener uses $\bar{h}$, the acceptance probability conditional on the state being y is $\sum_{x \in A} \zeta^*(y, x)[\sum_{m \in \sigma(x) \cap \sigma(y)} h(x, m) h^\circ(x, m, m) + \sum_{m \in \sigma(x) \setminus \sigma(y)} h(x, m) h^\circ(x, m, m^0) + \sum_{m' \in M \setminus \sigma(x)} h(x, m') h^\circ(x, m', m^0)]$. If the listener uses $\bar{g}$, the acceptance probability conditional on the state being y is $\sum_{x \in A} \zeta^*(y, x)[\sum_{m \in \sigma(x) \cap \sigma(y)} h(x, m) h^\circ(x, m, m) + (|\sigma(x) \cap \sigma(y)|/|\sigma(x)|) \sum_{m' \in M \setminus \sigma(x)} h(x, m') h^\circ(x, m', m^0)]$. When $y \in R$, the acceptance probability is (weakly) smaller under $\bar{g}$ than under $\bar{h}$.

[34] A claim that the state belongs to R leads to certain rejection.

The listener's problem of choosing a straightforward best reply to ζ^* can be represented as that of choosing $(g(x,m): x \in A, m \in \sigma(x))$ to minimize

$$\sum_{x \in A} \widehat{p}_x \left[\sum_{m \in \sigma(x)} g(x,m) \frac{\delta_{x,m}}{\widehat{p}_x} + \left(1 - \sum_{m \in \sigma(x)} g(x,m)\right) \frac{p_x}{\widehat{p}_x} \right]$$

subject to $\sum_{m \in \sigma(x)} g(x,m) \leq 1 \; \forall x \in A$ and $g(x,m) \geq 0 \; \forall x \in A, \forall m \in \sigma(x)$. The optimality conditions for the above optimization problem are

$$\sum_{m' \in M} g(x,m') < 1 \quad \Rightarrow \quad p_x \leq \delta_{x,m} \quad \forall x \in A, \forall m \in \sigma(x) \tag{42}$$

$$g(x,m) > 0 \quad \Rightarrow \quad \delta_{x,m} \leq p_x \quad \forall x \in A, \forall m \in \sigma(x) \tag{43}$$

$$g(x,m) > 0 \quad \Rightarrow \quad \delta_{x,m} \leq \delta_{x,m'} \quad \forall x \in A, \forall m, m' \in \sigma(x). \tag{44}$$

To establish that $\bar{g}^*$ is a best reply to $\bar{\zeta}^*$, it is sufficient to establish that g^* satisfies (42)–(44).

Constraints (16)–(19) and (20) imply that for all $x \in A$ and $m \in \sigma(x)$,

$$\begin{aligned}
\varphi^*(s,x) &\leq \sum_{y \in R: m \in \sigma(y)} \varphi^*(x,y) = \sum_{y \in R: m \in \sigma(y), \varphi^*(y,t) > 0} \frac{\varphi^*(x,y)}{\varphi^*(y,t)} \varphi^*(y,t) \\
&\leq \sum_{y \in R: m \in \sigma(y), \varphi^*(y,t) > 0} \frac{\varphi^*(x,y)}{\varphi^*(y,t)} p_y = \sum_{y \in R: m \in \sigma(y), \varphi^*(y,t) > 0} \zeta^*(y,x) p_y \\
&\leq \sum_{y \in R: m \in \sigma(y)} \zeta^*(y,x) p_y.
\end{aligned} \tag{45}$$

Consider any $x \in A$ such that $\sum_{m' \in \sigma(x)} g^*(x,m') < 1$. Then (4) and (8) imply $\mu_x^* = 1 - \sum_{m' \in \sigma(x)} \beta^*(x,m') > 0$. So complementary slackness implies $p_x = \varphi^*(s,x)$. Condition (45) then implies that $p_x \leq \sum_{y \in R: m \in \sigma(y)} \zeta^*(y,x) p_y$ for all $m \in \sigma(x)$. So g^* satisfies (42).

Consider $x \in A$ and $m \in \sigma(x)$ with $g^*(x,m) > 0$. We argue that in this case, all three inequalities in (45) become equalities. Equation (8) implies that $\beta^*(x,m) > 0$. So complementary slackness implies that the first inequality becomes an equality. For any $y \in R$ with $m \in \sigma(y)$, (5) implies $\mu_y^* > 0$. So complementary slackness implies $\varphi^*(y,t) = p_y$, implying that the second inequality becomes an equality. Choose any $y \in R$ with $\varphi^*(y,t) = 0$. Because $p_y > 0$, it follows that $\varphi^*(y,t) < p_y$. Complementary slackness implies that $\mu_y^* = 0$, which implies via (5) that because $\beta^*(x,m) > 0$, $m \notin \sigma(y)$. So the third inequality becomes an equality. To summarize, $\forall x \in A, \forall m \in \sigma(x), g^*(x,m) > 0 \Rightarrow \varphi^*(s,x) = \sum_{y \in R: m \in \sigma(y)} \zeta^*(y,x) p_y$. Constraint (15) now implies that g^* satisfies (43). For all $x \in A$ and $m' \in \sigma(x)$, (45) says that $\varphi^*(s,x) \leq \sum_{y \in R: m' \in \sigma(y)} \zeta^*(y,x) p_y$, implying that g^* satisfies (44). So $\bar{g}^*$ is a best reply to $\bar{\zeta}^*$. The fact that $\bar{\zeta}^*$ is a best reply to $\bar{g}^*$ implies that the equilibrium $(\bar{\zeta}^*, \bar{g}^*)$ induces the same error probability as the optimal rule g^* at every state.

Strengthening the Bayesian Nash equilibrium to a sequential equilibrium. Finally we modify $\bar{\zeta}^*$ and $\bar{g}^*$ so that the above Bayesian Nash equilibrium becomes a sequential equilibrium. Define $M_A := \{m \in M : \forall x \in X, m \in \sigma(x) \Rightarrow x \in A\}$, $A^* :=$

$\{x \in A : \sigma(x) \cap M_A \neq \emptyset\}$. For each $x \in A^*$, choose $m^*(x) \in \sigma(x) \cap M_A$. The speaker strategy $\bar{\zeta}^{**}$ agrees with $\bar{\zeta}^*$ with two exceptions: (i) Suppose $x \in X \setminus A^*$ and $\hat{x} \in X$, $m \in M$ satisfy $\sum_{x' \in X} p_{x'} \zeta^*(x', \hat{x}) g^*(\hat{x}, m) = 0$. Then if the state is x, the speaker presented the cheap talk report $\hat{x}$, and the listener requested m, the speaker will present m^0. (ii) Suppose $x \in A^*$ and $\hat{x} \in X$, $m \in M$ satisfy $p_x \zeta^*(x, \hat{x}) g^*(\hat{x}, m) = 0$. Then if the state is x, the speaker presented cheap talk report $\hat{x}$, and the listener requested m, the speaker will present $m(x)$. The listener strategy $\bar{g}^{**}$ agrees with $\bar{g}^*$ with two exceptions: (i′) Suppose $\hat{x} \in X$, $m \in M$ satisfy $\sum_{x' \in X} p_{x'} \zeta^*(x', \hat{x}) g^*(\hat{x}, m) \zeta^\circ(x', \hat{x}, m, m') = 0$. Then if the speaker presented cheap talk report $\hat{x}$, the listener requested message m, and the speaker presented some message $m' \in M_A$, then the listener will accept the speaker. (ii′) Suppose $\hat{x} \in A \setminus A^*$ and $m \in M$ satisfy $g^*(\hat{x}, m) = 0$. Then if the speaker presented cheap talk report $\hat{x}$, the listener requested message m, and the speaker presented message m^0, the listener will accept the speaker's request exactly if $p_{\hat{x}} > \sum_{y \in R} \zeta^*(y, \hat{x}) p_y$.

The strategy profiles $(\bar{g}^{**}, \bar{\zeta}^{**})$ and $(\bar{g}^*, \bar{\zeta}^*)$ differ only on zero probability histories, so $(\bar{g}^{**}, \bar{\zeta}^{**})$ induces the same error probability as the optimal rule g^* at every state. We now verify that $\bar{g}^{**}$ and $\bar{\zeta}^{**}$ are mutual best replies. Because $(\bar{g}^{**}, \bar{\zeta}^{**})$ and $(\bar{g}^*, \bar{\zeta}^*)$ differ only on zero probability histories and (g^*, ζ^*) is a Bayesian Nash equilibrium, to show that $\bar{g}^{**}$ is a best reply to $\bar{\zeta}^{**}$, it is sufficient to show that conditional on any report $\hat{x} \in A$, the listener has no incentive to request any message m with $g^*(\hat{x}, m) = 0$. If $\hat{x} \in A^*$, then the facts that $(\bar{g}^{**}, \bar{\zeta}^{**})$ implements an optimal rule and $\bar{\zeta}^{**}$ entails truthful reports on A imply that the listener is already achieving an error probability of zero conditional on report $\hat{x}$, and so has no incentive to deviate. If $\hat{x} \in A \setminus A^*$, then requesting a message m with $g^*(\hat{x}, m) = 0$ against $\bar{\zeta}^{**}$ would lead to the same outcome as either rejecting the speaker without requesting any messages against $\bar{\zeta}^*$ or requesting m^0 and then continuing according to $\bar{g}^*$ against $\bar{\zeta}^*$. Since these deviations were not attractive in the equilibrium $(\bar{g}^*, \bar{\zeta}^*)$, they cannot benefit the listener given the strategy profile $(\bar{g}^{**}, \bar{\zeta}^{**})$. So $\bar{g}^{**}$ is a best reply to $\bar{\zeta}^{**}$. The speaker can only cause $\bar{g}^*$ and $\bar{g}^{**}$ to differ by presenting $m \in M_A$, which is only possible in A^*. As explained above, in A^*, $\bar{g}^{**}$ avoids error and thus already accepts the speaker with probability 1 if the speaker uses $\bar{\zeta}^{**}$. It follows that $\bar{\zeta}^{**}$ is a best reply to $\bar{g}^{**}$.

We now argue that in $(\bar{g}^{**}, \bar{\zeta}^{**})$, the speaker's strategy is sequentially rational off the equilibrium path. If $x \in A^*$, this follows from the fact that conditional on any nonterminal history where the speaker moves, the probability of acceptance is 1. If $x \in X \setminus A^*$, then at all nonterminal off-equilibrium path histories where the speaker moves, the speaker has previously made a cheap talk claim $\hat{x}$ and the listener has requested a message m. Following such histories, according to $\bar{\zeta}^{**}$, either every available message will lead to rejection or the speaker will present the only available message—either m or m^0, depending on the history—which will lead to acceptance. This establishes sequential rationality for the speaker.

We can represent $\bar{\zeta}^{**}$ as a pair (ζ^*, ζ°), where ζ^* is the speaker's reporting strategy and $\zeta^\circ(x, \hat{x}, m, m')$ is the probability that the speaker presents m if the true state is x, the speaker claimed that the state is $\hat{x}$, and the listener requested message m'. For each sufficiently small $\epsilon > 0$, define a (totally mixed) speaker strategy $\bar{\zeta}_\epsilon := (\zeta_\epsilon, \zeta_\epsilon^\circ)$ by $\zeta_\epsilon(x, \hat{x}) := \zeta^*(x, \hat{x}) - \lambda^x$ (λ^x defined below) if $\zeta^*(x, \hat{x}) > 0$, by $\zeta_\epsilon(x, \hat{x}) := \epsilon$ if $\zeta^*(x, \hat{x}) = 0$ and $x \in R$, and by $\zeta_\epsilon(x,\hat{x}) := \epsilon^2$ otherwise. Define $\zeta^\circ_\epsilon(x,\hat{x},m,m') := \zeta^\circ(x,\hat{x},m,m') - \tau^{x,\hat{x},m}$ ($\tau^{x,\hat{x},m}$ defined below) if $\zeta^\circ(x,\hat{x},m,m') > 0$, define $\zeta^\circ_\epsilon(x,\hat{x},m,m') := \epsilon$ if $\zeta^\circ(x,\hat{x},m,m') = 0$, $m' \in \sigma(x)$, and $x \in R$ and define $\zeta^\circ_\epsilon(x,\hat{x},m,m') = \epsilon^2$ if $\zeta^\circ(x,\hat{x},m,m') = 0$, $m' \in \sigma(x)$, and $x \in A$. The values of λ^x and $\tau^{x,\hat{x},m}$ are chosen so that $\sum_{\hat{x}\in X}\zeta_\epsilon(x,\hat{x}) = 1$ and $\sum_{m'\in\sigma(x)}\zeta^\circ_\epsilon(x,\hat{x},m,m') = 1$, and $\bar{\zeta}^\circ_\epsilon \to \bar{\zeta}^{**}$ as $\epsilon \to 0$. Let b be the limiting listener beliefs about the state (as a function of the listener's information set) induced by $\bar{\zeta}^\circ_\epsilon$ as $\epsilon \to 0$ and *any* totally mixed listener strategy.[35] There are two kinds of off-equilibrium listener information sets: (i) those where the listener knows that the speaker took some action, which occurs with zero probability according to $\bar{\zeta}^{**}$ given the listener's past actions, and (ii) the off-equilibrium information sets not in the first category. In case (i), unless the speaker has presented a message $m \in M_A$, according to b, the listener assigns probability 1 to the state being in R. Following such histories, the listener will always reject the speaker according to $\bar{g}^{**}$, a best reply to his beliefs. In histories of type (i) where the speaker presents $m \in M_A$, the listener will assign probability 1 to the state belonging to A according to b and will accept the speaker, a best reply. Finally following histories of type (ii), $\bar{g}^{**}$ is constructed so that the listener will reject the speaker exactly if according to b, he assigns probability of at most $\frac{1}{2}$ to the state belonging to A, again, a best reply. This establishes that $\bar{g}^*$ is sequentially rational against beliefs b, completing the proof. □

Proof of Theorem 6.

Proof of part (i) of Theorem 6. Constraint (17) is equivalent to

$$\varphi(s,x) \le \min_{m\in\sigma(x)} \sum_{y\in R: m\in\sigma(y)} \varphi(x,y) \quad \forall x \in A. \tag{46}$$

Assume foresight. Definition 1 and (19) imply $m_x \in \arg\min_{m\in\sigma(x)} \sum_{y\in R: m\in\sigma(y)} \varphi(x,y)$ $\forall x \in A$. Definition 1 implies $m_x \in \sigma(x) \cap \sigma(y) \Leftrightarrow \sigma(x) \subseteq \sigma(y)$ $\forall x \in A$, $\forall y \in R$. So (46) simplifies to $\varphi(s,x) \le \sum_{y:\sigma(x)\subseteq\sigma(y)} \varphi(x,y)$ $\forall x \in A$. At any optimum of Program D,

$$\varphi(s,x) = \sum_{y:\sigma(x)\subseteq\sigma(y)} \varphi(x,y) \quad \forall x \in A. \tag{47}$$

Otherwise, we could increase $\varphi(s,x)$ without violating any constraints, increasing the objective. Replacing (17) by (47), we may assume that $\varphi(x,y) = 0$ whenever $\sigma(x) \nsubseteq \sigma(y)$; otherwise, we could reduce $\varphi(y,t)$ by $\varphi(x,y)$ and set $\varphi(x,y) = 0$, attaining a new feasible solution with the same objective value. So (18) becomes

$$\sum_{x:\sigma(x)\subseteq\sigma(y)} \varphi(x,y) = \varphi(y,t) \quad \forall y \in R. \tag{48}$$

Given edges (22), (47) and (48) are equivalent to (25) and (14) is equivalent to (24). Equation (23) implies equivalence of (15) and (16) with (26). Constraint (19) and equations (47) and (48) imply (27). So Program D reduces to Max Flow. ◁

[35]The beliefs b are independent of the listener's totally mixed strategy.

Proof of part (ii) of Theorem 6.

Lemma 1. *Adding or removing constraints (31) has no affect on the value of Min Cut.*

Proof. For any optimal solution (δ, γ) to Min Cut without (31), define $V_- := \{v : \gamma_v < 0\}$, $V_+ := \{v : \gamma_v > 1\}$, and $V_0 := V \setminus (V_- \cup V_+)$. Define a new solution (δ, γ'), where $\gamma'_v := 0\ \forall v \in V_-$ and $\gamma'_v = 1\ \forall v \in V_+$, and otherwise (δ, γ') coincides with (δ, γ). The solution (δ, γ') satisfies (30)–(33) and in particular (31). We verify that (δ, γ') satisfies (31) using optimality (and hence feasibility) of (δ, γ). First choose $(s, x) \in E_1$. If $x \in V_0$, (31) is immediate. If $x \in V_-$, $\delta(s, x) \geq \gamma_s - \gamma_x = 1 - \gamma_x > 1 = 1 - \gamma'_x = \gamma'_s - \gamma'_x$. If $x \in V_+$, then $\delta(s, x) \geq 0 = 1 - 1 = \gamma'_s - \gamma'_x$. Choose $(y, t) \in E_3$. If $y \in V_0$, (31) is immediate. If $y \in V_-$, then $\delta(y, t) \geq 0 = 0 - 0 = \gamma'_t - \gamma'_y$. If $y \in V_+$, then $\delta(y, t) \geq \gamma_y - \gamma_t = \gamma_y \geq \gamma'_y = \gamma'_y - \gamma'_t$. Choose $(x, y) \in E_2$. Optimality of (δ, γ) implies $\delta(x, y) = 0$. So (31) reduces to $\gamma'_y \geq \gamma'_x$. Since $\gamma_y \geq \gamma_x$, (x, y) must belong to one of the following sets: $V_- \times V_-$, $V_- \times V_0$, $V_- \times V_+$, $V_0 \times V_0$, $V_0 \times V_+$, or $V_+ \times V_+$. If $(x, y) \in V_0 \times V_0$, then $\gamma'_y = \gamma_y \geq \gamma_x = \gamma'_x$. In all other cases, (31) follows from the definition of γ'. Since (δ, γ) and (δ, γ') have the same objective value, we are done. ◁

Lemma 1 justifies treating (29)–(33) as the dual of Max Flow below.

Lemma 2. *At any optimal solution to Min Cut, (i) $\delta(s, x) = 1 - \gamma_x\ \forall (s, x) \in E_1$, (ii) $\delta(x, y) = 0\ \forall (x, y) \in E_2$, and (iii) $\delta(y, t) = \gamma_y\ \forall (y, t) \in E_3$.*

Proof. Min Cut has a finite value.[36] This, along with (23), implies (ii). Equations (29) and (32) imply $\delta(s, x) \geq \gamma_s - \gamma_x = 1 - \gamma_x\ \forall x \in A$. If the inequality is strict, (31) implies we can reduce $\delta(s, x)$ without violating any constraints, reducing the objective, implying (i). Similarly, for any $y \in R$, $\delta(y, t) \geq \gamma_y - \gamma_t = \gamma_y$, and if the inequality is strict, we can reduce $\delta(y, t)$, implying (iii). ◁

Fix an optimum (δ, γ) in Min Cut. Define (μ, β) by (34). We argue that (μ, β) is feasible in Program P0. Equation (34) implies $\forall x \in A$, $1 - \mu_x = 1 - (1 - \gamma_x) = \gamma_x = \sum_{m \in \sigma(x)} \beta(x, m)$. Constraint (29) and equation (34), and (ii) of Lemma 2 imply $\forall x \in A$, $\forall y \in R$, $\sigma(x) \subseteq \sigma(y) \Rightarrow \mu_y = \gamma_y \geq \gamma_x - \delta(x, y) = \gamma_x = \sum_{m \in \sigma(x) \cap \sigma(y)} \beta(x, m)$. Constraint (31) and equation (34), and foresight imply $\forall x \in A$, $\forall y \in R$, $\sigma(x) \nsubseteq \sigma(y) \Rightarrow \mu_y \geq 0 = \sum_{m \in \sigma(x) \cap \sigma(y)} \beta(x, m)$. Given the above relations, it is straightforward to verify that (μ, β) is feasible in Program P0. Equation (34), Lemma 2, (23), and Max Flow–Min Cut duality imply $\sum_{z \in X} p_z \mu_z = \sum_{x \in A} p_x (1 - \gamma_x) + \sum_{y \in R} p_y \gamma_y = \sum_{(v, w) \in E} \delta(v, w) c(v, w) = \sum_{x \in A} \varphi(s, x)$ for any optimum φ of Max Flow. Part (i) of Theorem 6 and duality of Programs P0 and D now imply optimality of (μ, β). ◁

□

Solution to Example 4. The maximum flow φ^* for Example 4 is given by Table 2. The table is split into three panels. The top, center, and bottom panels describe the flow

[36] Zero is a lower bound on the value; a feasible solution with finite value is $\delta(s, x) = 1\ \forall (s, x) \in E_1$; $\delta(v, w) = 0\ \forall (v, w) \notin E_1$; $\gamma_s = 1$; $\gamma_v = 0\ \forall v \neq s$.

(v, w)	$\varphi^*(v, w)$	$c(v, w)$
$(s, 2)$	$\frac{2}{16}$	$\frac{2}{16}$
$(s, 1)$	$\frac{2}{16}$	$\frac{3}{16}$
$(s, 4)$	0	$\frac{1}{16}$
$(s, 5)$	$\frac{2}{16}$	$\frac{2}{16}$
$(2, 6)$	$\frac{2}{16}$	∞
$(2, 7)$	0	∞
$(2, 8)$	0	∞
$(1, 7)$	$\frac{1}{16}$	∞
$(1, 8)$	$\frac{1}{16}$	∞
$(5, 8)$	$\frac{2}{16}$	∞
$(3, t)$	0	$\frac{1}{16}$
$(6, t)$	$\frac{2}{16}$	$\frac{3}{16}$
$(7, t)$	$\frac{1}{16}$	$\frac{1}{16}$
$(8, t)$	$\frac{3}{16}$	$\frac{3}{16}$

TABLE 2. A maximum flow for Example 4.

φ^* on edges in E_1, E_2, and E_3, respectively (see (22)). In each panel, an entry in the left column names an edge (v, w), the corresponding entry in the center column gives the flow $\varphi^*(v, w)$, and the entry in the right column gives the capacity $c(v, w)$.

Figure 4 depicts the maximum flow.[37] All edges point from left to right, in the direction away from s and toward t, but arrows have been omitted from the figure. The thick black lines depict positive flow along an edge; edges without thick black lines have zero flow. If an edge e consists of two parallel lines, this means that the flow along the edge is less than the capacity (i.e., $\varphi^*(e) < c(e)$), whereas if it consists of only one line, this means that the flow is equal to the capacity. The residual capacity of edge e is $c(e) - \varphi^*(e)$. The residual capacity of edges that connect the source to A represents the net benefit of acceptance or, more precisely, the probability of accepting the speaker minus the probability of incorrectly accepting the speaker. Similarly, the residual capacity of edges that connect R to the sink represents the net benefit of rejection.

A minimum cut Z^* is $\{s, 1, 4, 5, 7, 8\}$. This is depicted in the figure as the set of vertices underneath the curve. The capacity of this cut is the sum of capacities of edges that cross from Z^* to its complement; in the figure, this is the sum of capacities of edges that cross the curve from below, moving in a northeast direction. So the capacity of the minimum cut Z^* is $\frac{6}{16}$, which is the same as the value of the maximum flow (i.e., the quantity of flow exiting the source s). Cut Z^* consists of the set of vertices reachable from the source on a path that travels forward along edges with positive residual capacity or backward along edges with positive flow. Corollary 2 implies that there is an optimal persuasion rule that accepts exactly messages $\{m_1, m_4, m_5\}$ and thereby accepts

[37]The positions of vertices 1 and 2 have been interchanged in comparison to Figure 3, but this difference is merely cosmetic. The two figures represent exactly the same network because they contain the same vertices and edges.

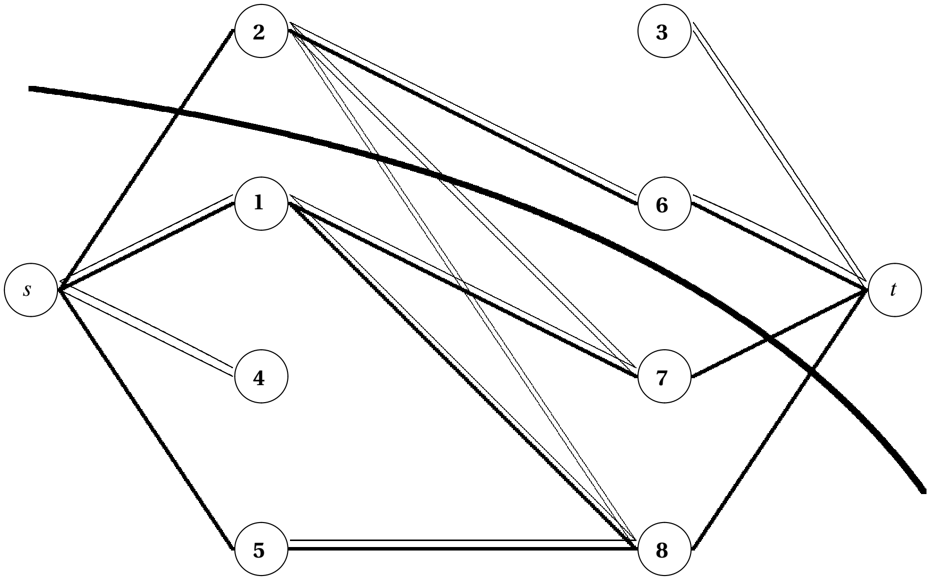

FIGURE 4. A maximum flow and minimum cut for Example 4.

the speaker exactly in the states under the curve $\{1, 4, 5, 7, 8\}$. Corollary 2 also shows how to derive the speaker strategy in the credible implementation. At states $x \in A$, the speaker reports truthfully. At states $y \in R$, the speaker strategy is derived via the maximum flow. For example, at state $8 \in R$, the speaker claims that the state is 1 and 5 with probabilities proportional to the flows on edges $(1, 8)$ and $(5, 8)$: At 8, the speaker claims the state is 1 with probability $\varphi^*(1, 8)/\varphi^*(8, t) = \frac{1}{3}$ and that the state is 5 with probability $\varphi^*(5, 8)/\varphi^*(8, t) = \frac{2}{3}$. Finally, Theorem 6 tells us that the total error probability at the optimal rule is equal to $\frac{6}{16}$, the value of the maximum flow.

PROOF OF COROLLARY 3. Let Z^* be the union of all minimum cuts. (See the definition of a minimum cut in the paragraph preceding Corollary 2.) Due to the lattice structure of the set of minimum cuts, Z^* is itself a minimum cut (Ford and Fulkerson 1956, Topkis 1998). Define $f^*(m) := 1$ if $\{x \in X : m \in \sigma(x)\} \subseteq Z^*$ and define $f^*(m) := 0$ otherwise. The rule f^* accepts the speaker precisely at states in Z^*. Theorem 6 and Corollary 1 imply that f^* is an optimal persuasion rule. By construction, f^* is less difficult than any other optimal deterministic static rule and f^* is symmetric. Assume for contradiction that there is an optimal static rule f' and message m' such that $f^*(m') = 0 < f'(m')$. Define (δ', γ') so that $\gamma'_x := \alpha(f', x)$ and δ' satisfies conditions (i)–(iii) in Lemma 2. Theorem 6 and optimality of persuasion rule f' imply that (δ', γ') is an optimal solution to Min Cut. Because the constraint matrix of Min Cut is totally unimodular, the set of optimal solutions to Min Cut is a polytope with integral extreme points, and (δ', γ') must be a convex combination of these integral extreme points. For each such integral optimum (δ'', γ''), there exists a corresponding optimal rule defined by $f''(m) := 1$ if $\{x \in X : m \in \sigma(x)\} \subseteq \{x \in X : \gamma_x = 1\}$ and by $f''(m) := 0$ otherwise. At least one of these deterministic optimal rules must be such that $f''(m') = 1$, contradicting the fact that f^*

is less difficult than all deterministic optimal rules. It follows that f^* is less difficult than all static optimal rules. □

Proof of part (i) of Theorem 7. Let U be a set and let $\mathcal{S}$ be a family of subsets of U such that $\bigcup \mathcal{S} = U$. The *set cover problem* is the problem of finding a minimal cardinality subset $\mathcal{T}$ of $\mathcal{S}$ such that $\bigcup \mathcal{T} = U$. This problem is known to be NP-hard. We will prove that the static persuasion problem (without foresight) is NP-hard by reduction from the set cover problem. Consider an instance of the set cover problem $(U, \mathcal{S})$. For each $S \in \mathcal{S}$, construct a state x_S and let $X = U \cup \{x_S : S \in \mathcal{S}\}$. Let $A = U$ and $R = \{x_S : S \in \mathcal{S}\}$. For each $x \in A$, let $p_x := 2|R|/[(2|A|+1)|R|]$ and for each $y \in R$, let $p_y := 1/[(2|A|+1)|R|]$. For each $S \in \mathcal{S}$, let there be a message m_S and define σ so that $m_S \in \sigma(x) \Leftrightarrow x \in S \cup \{x_S\}$. We know that there is an optimal static persuasion rule that is deterministic and that any deterministic optimal static persuasion rule can be associated with the set K of messages that it accepts. Given the specification of the above persuasion problem, it is clear that any deterministic optimal persuasion rule must accept all states in $A = U$ or, more precisely, must accept some message available at x for all $x \in A$. The goal is then to do this while accepting the minimum number of states in R, which amounts to finding a minimal cardinality collection $K \subseteq M$ such that for all $x \in A$, $K \cap \sigma(x) \neq \emptyset$. This is equivalent to finding a minimal cardinality subset $\mathcal{T}$ of $\mathcal{S}$ that covers U (i.e., $\bigcup \mathcal{T} = U$), which is the set cover problem. □

References

Ahuja, Ravindra K., Thomas L. Magnanti, and James B. Orlin (1993), *Network Flows*. Prentice Hall, Englewood Cliffs, New Jersey. [101, 115, 118]

Aumann, Robert J. and Sergiu Hart (2003), "Long cheap talk." *Econometrica*, 71, 1619–1660. [101]

Ben-Porath, Elchanan and Barton L. Lipman (2012), "Implementation with partial provability." *Journal of Economic Theory*, 147, 1689–1724. [113]

Bull, Jesse and Joel Watson (2004), "Evidence disclosure and verifiability." *Journal of Economic Theory*, 118, 1–31. [113]

Bull, Jesse and Joel Watson (2007), "Hard evidence and mechanism design." *Games and Economic Behavior*, 58, 75–93. [102, 105, 113, 118]

Chen, Xi and Xiaotie Deng (2006), "Settling the complexity of two-player Nash equilibrium." In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 261–272, IEEE Computer Society, Los Alamitos. [121]

Chen, Ying and Wojciech Olszewski (2011), "Effective persuasion." Unpublished paper, Arizona State University and Northwestern University. [101]

Conitzer, Vincent and Tuomas Sandholm (2008), "New complexity results about Nash equilibria." *Games and Economic Behavior*, 63, 621–641. [121]

Crawford, Vincent P. and Joel Sobel (1982), “Strategic information transmission.” *Econometrica*, 50, 1431–1451. [101]

Daskalakis, Constantinos, Paul W. Goldberg, and Christos H. Papadimitriou (2009), “The complexity of computing a Nash equilibrium.” *SIAM Journal on Computing*, 39, 195–259. [121]

Deneckere, Raymond and Sergei Severinov (2008), “Mechanism design with partial state verifiability.” *Games and Economic Behavior*, 64, 487–513. [113]

Dziuda, Wioletta (2011), “Strategic argumentation.” *Journal of Economic Theory*, 146, 1362–1397. [101, 113]

Fishman, Michael J. and Kathleen M. Hagerty (1990), “The optimal amount of discretion to allow in disclosure.” *Quarterly Journal of Economics*, 105, 427–444. [119]

Ford, Lester R. and Delbert R. Fulkerson (1956), “Maximal flow through a network.” *Canadian Journal of Mathematics*, 8, 399–404. [101, 120, 133]

Forges, Françoise and Frédéric Koessler (2005), “Communication equilibria with partially verifiable types.” *Journal of Mathematical Economics*, 41, 793–811. [113]

Gilboa, Izhak and Eitan Zemel (1989), “Nash and correlated equilibria: Some complexity considerations.” *Games and Economic Behavior*, 1, 80–93. [121]

Glazer, Jacob and Ariel Rubinstein (2001), “Debates and decisions: On a rationale of argumentation rules.” *Games and Economic Behavior*, 36, 158–173. [101]

Glazer, Jacob and Ariel Rubinstein (2004), “On optimal rules of persuasion.” *Econometrica*, 72, 1715–1736. [101, 102, 109, 110, 112]

Glazer, Jacob and Ariel Rubinstein (2006), “A study in the pragmatics of persuasion: A game theoretical approach.” *Theoretical Economics*, 1, 395–410. [101, 102, 107, 109, 110, 112, 119, 123]

Goldreich, Oded (2008), *Computational Complexity: A Conceptual Perspective*. Cambridge University Press, Cambridge. [120]

Goltsman, Maria, Johannes Hörner, Gregory Pavlov, and Francesco Squintani (2009), “Mediation, arbitration and negotiation.” *Journal of Economic Theory*, 144, 1397–1420. [101]

Green, Jerry R. and Jean-Jacques Laffont (1986), “Partially verifiable information and mechanism design.” *Review of Economic Studies*, 53, 447–456. [102, 113]

Kartik, Navin and Olivier Tercieux (2012), “Implementation with evidence.” *Theoretical Economics*, 7, 323–355. [113]

Krishna, Vijay and John Morgan (2004), “The art of conversation: Eliciting information from experts through multi-stage communication.” *Journal of Economic Theory*, 117, 147–179. [101]

Lipman, Barton L. and Duane J. Seppi (1995), "Robust inference in communication games with partial provability." *Journal of Economic Theory*, 66, 370–405. [102, 113]

Milgrom, Pual R. (1981), "Good news and bad news: Representation theorems and applications." *Bell Journal of Economics*, 12, 380–391. [113]

Milgrom, Paul R. and John Roberts (1986), "Relying on the information of interested parties." *RAND Journal of Economics*, 17, 18–32. [101, 113]

Osborne, Martin J. and Ariel Rubinstein (1994), *A Course in Game Theory*. MIT Press, Cambridge, Massachusetts. [105]

Picard, Jean-Claude (1976), "Maximal closure of a graph and applications to combinatorial problems." *Management Science*, 22, 1268–1272. [115]

Sher, Itai (2008), "Persuasion and limited communication." Minnesota Economic Research Report 2008-2, University of Minnesota. [101, 112]

Sher, Itai (2011), "Credibility and determinism in a game of persuasion." *Games and Economic Behavior*, 71, 409–419. [110]

Sher, Itai (2013), "Persuasion and the minimax theorem." Unpublished paper, University of Minnesota. [121]

Sher, Itai and Rakesh Vohra (2013), "Price discrimination through communication." Unpublished paper, University of Minnesota and Northwestern University. [121]

Shin, Hyun Song (1994a), "The burden of proof in a game of persuasion." *Journal of Economic Theory*, 64, 253–264. [101, 113]

Shin, Hyun Song (1994b), "News management and the value of firms." *RAND Journal of Economics*, 25, 58–71. [113]

Shin, Hyun Song (2003), "Disclosure and asset returns." *Econometrica*, 71, 105–133. [113]

Singh, Nirvikar and Donald Wittman (2001), "Implementation with partial verification." *Review of Economic Design*, 6, 63–84. [113]

Topkis, Donald M. (1998), *Supermodularity and Complementarity*. Princeton University Press, Princeton. [133]

Vohra, Rakesh V. (2011), *Mechanism Design: A Linear Programming Approach*. Cambridge University Press, Cambridge. [121]

Submitted 2009-11-23. Final version accepted 2012-11-3. Available online 2012-11-3.