Credibility and Determinism in a Game of Persuasion

Itai Sher*
University of Minnesota

January 2, 2008

Abstract

This paper continues a project initiated by Glazer and Rubinstein (2001, 2004, 2006) of studying optimal rules of persuasion. A speaker attempts to persuade a listener to take an action by presenting evidence. I extend results establishing credibility and determinism of optimal persuasion rules from a model with two actions to one with many actions. In the latter case, additional assumptions on the utility functions are needed. I introduce a notion of quasi-determinism, which involves a limited form of randomization, and show that there always exists an optimal persuasion rule with this property. I also present an assumption weaker than optimality which is sufficient for credibility.

1 Introduction

This paper continues a project initiated by Glazer and Rubinstein (2001, 2004, 2006) of studying optimal rules of persuasion. These papers involve a speaker or speakers attempting to persuade a listener to take an action by presenting evidence. A choice of persuasion rule is a commitment to respond to evidence in a specific way.

I extend results establishing credibility and determinism of optimal persuasion rules from a model with two actions to one with many actions. Credibility means that there is no

---

*I am very grateful to Eddie Dekel for many valuable comments and advice. I would like to thank David Austen-Smith, Marciano Siniscalchi, Asher Wolinsky, Ariel Rubinstein, Jesse Bull, Elchanan Ben-Porath, David Rahman, and Tomasz Strzalecki. I am also grateful to seminar participants at The University of British Columbia, University of Illinois at Urbana-Champaign, Princeton University, University of Minnesota, University of Rochester, Penn State, and Lund University for their comments. All errors are mine.

commitment problem involved in implementing the optimal rule, and determinism means that the rule involves no randomization.

Consideration of many actions highlights the importance of properties which are irrelevant with two actions. The analogs of the two action results are only true under additional assumptions. These additional assumptions shed light on the persuasion problem.

With two actions, there always exist deterministic optimal rules and all optimal rules are credible. Neither of these results are true with many actions. In particular, Glazer and Rubinstein (2006) provided a counterexample to the credibility result with three actions. One might conclude that binary actions are essential to the credibility result. A sharper analysis shows that *concavity* is the essential property for credibility. More specifically, credibility holds when there exists a monotone function r of actions $A := \{1, \dots n\}$–ordered by the speaker's preference–such that the listener's utility function is always the restriction of a concave transformation of r to A. When A contains two elements, this is vacuously true, hence credibility always holds.

A similar property is essential for determinism, establishing a relationship between credibility and determinism which is invisible in the binary case. Moreover, I characterize the form which randomization may take in optimal persuasion rules. In particular, there always exist optimal persuasion rules which are *quasi-deterministic*, meaning (i) the speaker expected utility of any message is equal to the utility of some pure action, and (ii) each message leads to a probability distribution over at most two actions; these actions are not adjacent. With two actions, (i) is equivalent to determinism, as is (ii) (because of non-adjacency). Quasi-determinism generalizes determinism, and this provides a second generalization of the determinism result with two actions.

Another result sharpens the credibility result further. I define a graph on the set of deterministic persuasion rules. Two persuasion rules are neighbors in this graph if one can be derived from the other by choosing some action and some set of messages assigned to this action, and re-assigning them to an adjacent action. I show that to attain the credibility result, it is sufficient to assume only that a persuasion rules is superior to all of its neighbors on this graph, rather than that it is globally optimal.

The paper is organized as follows: Section 2 presents the model. Section 3 presents results on determinism. Section 4 deals with credibility. results on local optimality. Section 5 concludes. Some proofs are contained in an appendix.

2 The Model

There is a speaker and a listener. The speaker may be one of finitely many types in a set T. $p(t) > 0$ is the probability that the speaker is of type t.

There is a finite set $\mathbf{M}$ of messages. These are the messages that the speaker could potentially send. However, different types of speaker may send different messages. This may be because the hard evidence which is available depends on the state of the world. Let $M : T \to 2^{\mathbf{M}} \setminus \{\emptyset\}$. $M(t)$ is interpreted as the set of messages available to a type t speaker. I assume that $\mathbf{M} = \bigcup_{t \in T} M(t)$. In other words, $\mathbf{M}$ is the set of messages which are available to some type of speaker. The tuple $\mathcal{M} := (T, \mathbf{M}, M(\cdot))$ is a **message structure**.

The listener has a finite set of actions $A = \{1, \ldots, n\}$ with $n \geq 2$. a is **adjacent** to a' if $a = a' + 1$ or $a = a' - 1$. I sometimes use the notation $\underline{a} := 1, \overline{a} := n$. The speaker has a utility function $u : A \to \mathbb{R}$. I assume that $a < a' \Rightarrow u(a) < u(a')$. Thus, we can think of the actions as being ordered by the speaker's preference. Notice that the speaker's utility function does not depend on his type. This means that the listener knows what the speaker would like the listener to do.

The listener has a utility function $v : A \times T \to \mathbb{R}$. Thus, the listener's preferred action does depend on the speaker's private information. The underlying idea is that we are interested in persuasion in an environment in which the listener knows what the speaker would like the listener to do, but is uncertain about what the speaker can say.

2.1 Optimal Persuasion Rules

A persuasion rule is a rule which assigns to each message a probability distribution over actions. For a persuasion rule f, $f_{m,a}$ is the probability that action a is chosen if message m is sent. A persuasion rule can be thought of as an element f of $[0,1]^{\mathbf{M} \times A}$ satisfying $\sum_{a \in A} f_{m,a} = 1$ for all $m \in \mathbf{M}$. F is the set of all persuasion rules. An interesting subset of persuasion rules contains rules which respond to each message by selecting some action with probability one. Formally, a persuasion rule f is **deterministic** if for all m and a, $f_{m,a} \in \{0, 1\}$. D is the set of all deterministic persuasion rules. If $f \in D$, I write $f(m)$ for the unique action a such that $f_{m,a} = 1$. Then $f^{-1}(a) := \{m \in \mathbf{M} : f(m) = a\}$.

A speaker strategy is a function which assigns to each type a probability distribution over messages. For any type t, positive probability may only be assigned to messages available to that type, or in other words, to messages in $M(t)$. For a speaker strategy σ, $\sigma_{t,m}$ is the probability that type t assigns to message m under that strategy. We may think of a speaker strategy as an element σ of $[0,1]^{T \times \mathbf{M}}$ satisfying $\sum_{m \in M(t)} \sigma_{t,m} = 1$ and $m \notin M(t) \Rightarrow \sigma_{t,m} = 0$ for all $t \in T$.

The problem we are interested in is that of selecting an optimal persuasion rule. The listener is assumed to commit to a persuasion rule before the speaker sends his message. For any persuasion rule f, $B(f)$ is the set of speaker best replies to f. Formally, $B(f)$ is the set of σ's which maximize $\sum_{m \in \mathbf{M}} \sum_{a \in A} \sigma_{t,m} f_{m,a} u(a)$ for all $t \in T$. A persuasion rule f^* is **optimal** if it maximizes $\max_{\sigma \in B(f)} \sum_{t \in T} \sum_{m \in \mathbf{M}} \sum_{a \in A} \sigma_{t,m} f_{m,a} v(a,t) p(t)$. A persuasion rule is **deterministically optimal** if it is optimal when one restricts attention to persuasion rules in D. Notice that f enters into the objective in two ways, one of which is through $B(f)$. Implicitly, in determining an optimal persuasion rule, it is as though the listener could select the speaker's strategy as well, provided that it is a best reply to the selected rule.

2.2 Game without Commitment

In addition to the listener's problem when the listener can commit, we will also be interested in a game without commitment. In such a game, we may think of the listener's selection of a persuasion rule as being simultaneous with the speaker's selection of strategy, or equivalently, and perhaps more intuitively, we may think of the speaker as first selecting a message and the listener responding with an action. The solution concept is Bayesian Nash equilibrium. Formally, an **equilibrium** of the game without commitment is a pair (σ^*, f^*) such that $\sigma^* \in B(f^*)$, and for all m, f^* maximizes $\sum_{t \in T} \sum_{a \in A} \sigma^*_{t,m} f_{m,a} v(a,t) p(t)$. We do not consider randomization over persuasion rules because there always exists an equilibrium in which some persuasion rule is chosen with probability one. A persuasion rule f^* is **credible** if there exists a σ^* such that (i) (σ^*, f^*) is an equilibrium, and (ii) $\sigma^* \in \operatorname{argmax}_{\sigma \in B(f)} \sum_{t \in T} \sum_{m \in \mathbf{M}} \sum_{a \in A} \sigma_{t,m} f_{m,a} v(a,t) p(t)$. It is easy to see that whenever f is deterministic, it is sufficient to check condition (i), since in this case, every best response for the speaker induces the same utility for the listener. Thus a persuasion rule is credible if there is some equilibrium in the game without commitment which supports it, in which the listener, moreover, chooses the best response which is most favorable to the listener. The rationale for condition (ii) is that we are interested in the question of whether optimal persuasion rules are credible, and optimal persuasion rules are evaluated on the assumption that the listener may choose among speaker best responses; we want to be sure that both the persuasion rule and the speaker best response that the listener chooses are consistent with equilibrium in the game without commitment. In this paper, we do not consider sequential rationality off the equilibrium path because consideration of off-path sequential rationality would not substantively alter any conclusions.

3 Determinism

Glazer and Rubinstein (2006) established that in the case of two actions, there always exists an optimal deterministic persuasion rule. This is no longer true in the case of many actions.

Example 1 Suppose that there are two types t_1 and t_2 and three actions, and that $M(t_i) =$

$\{m_1, m_2\}$ for $i = 1, 2$. Suppose that the listener's utility function is given by:

v	a_1	a_2	a_3
t_1	1	0	1
t_2	0	1	0

where I have written a_i instead of i for the ith action. I assume moreover that the speaker's utility function is such that $u(a_i) = i$. In any optimal rule, the response to one message is action a_2 with probability 1, and the response to the other puts half the probability on a_1 and the other half on a_3. $\square$

In this section, I present two generalizations of the determinism result to many actions. The following concept is weaker than determinism:

Definition 1 *A persuasion rule f is **quasi-deterministic** if for all $m \in \mathbf{M}$:*

$$\sum_{a \in A} u(a) f_{m,a} \in \{u(a) : a \in A\} \tag{1}$$

$$|\{a \in A : f_{m,a} > 0\}| \leq 2 \tag{2}$$

As is the case with deterministic persuasion rules, given a quasi-deterministic persuasion rule, every message gives the speaker an expected utility equal to that of some pure action. Randomization is further limited in that each message leads to at most two actions with positive probability. Of course, when randomizing, the probabilities must be $\frac{u(a''')-u(a'')}{u(a''')-u(a')}$ and $\frac{u(a'')-u(a')}{u(a''')-u(a')}$ for some three actions $a' < a'' < a'''$, and then the expected utility is equal to $u(a'')$. Moreover, the two action a' and a''' receiving positive probability cannot be adjacent. Observe that determinism implies quasi-determinism, and with only two actions the two are equivalent.

Define $R(\sigma) := \{f \in F : \sigma \in B(f)\}$. $R(\sigma)$ is the set of persuasion rules which **rationalize** σ. $R(\sigma)$ is always nonempty because it always contains all persuasion rules which respond to all messages in the same way. Being defined by a set of linear inequalities, bounded, and nonempty, $R(\sigma)$ is always a polytope (i.e., the convex hull of a finite number of points).

Theorem 1 *The extreme points of $R(\sigma)$ are quasi-deterministic.*

Proof. In appendix.

Corollary 1 *There exists an optimal quasi-deterministic persuasion rule.*

Proof. Let f^* be an optimal persuasion rule, and let σ^* the speaker strategy which maximizes the listener's expected utility within $B(f^*)$. Then any persuasion rule which maximizes the listener's expected utility within $R(\sigma^*)$ on the assumption that the speaker will use σ^* is optimal. We can find a solution to the latter problem among the extreme points of $R(\sigma^*)$. So the conclusion follows from Theorem 1. $\square$

Corollary 2 *In the case of two actions, there is an optimal deterministic persuasion rule.*

Proof. This follows from Corollary 1 and the fact that with two actions quasi-determinism and determinism are equivalent. $\square$

As mentioned above, the last corollary was proven in a different way in Glazer and Rubinstein (2006). In fact, Corollary 1 is a generalization of Glazer and Rubinstein's theorem to many actions.

Using an argument similar to that in Corollary 1, it follows that:

Corollary 3 *For every equilibrium* (σ^*, f^*) *of the game without commitment, there exists quasi-deterministic* f^{**} *such that* (σ^*, f^{**}) *is an equilibrium. In the case of two actions,* f^{**} *is deterministic.*

It is obvious that the listener–but possibly not the speaker–attains the same utility in the two equilibria. This theorem shows that in the case of two actions, not only is it always possible to find an optimal deterministic persuasion rule, but in response to any equilibrium speaker strategy there is a deterministic equilibrium listener best response. An analogous fact involving quasi-determinism holds with many actions.

I now introduce an assumption which is important for both determinism and credibility.

Assumption 1 *For all* $t \in T$*, there exists a concave function* $c_t : \mathbb{R} \to \mathbb{R}$ *such that for all* $a \in A$, $v(a, t) = c_t(u(a))$.

Corollary 4 *Given Assumption 1, there exists an optimal deterministic persuasion rule.*

Proof. By Corollary 1, we can choose an optimal quasi-deterministic persuasion rule f^*, and let σ^* be the speaker strategy which maximizes the listener's utility in $B(f^*)$ given that the listener uses f^*. By quasi-determinism, for any message m such that f^* selects a nondegenerate distribution δ in response to m, there is an action a such that a with probability 1 gives the speaker the same utility as δ. Construct a new persuasion rule f^{**} which replaces all such δ's with the corresponding a's. Then $\sigma^* \in B(f^{**})$. Moreover, by Jensen's inequality, the listener is better off. $\square$

Since every function on a two element set is the restriction of some concave function to that set, Corollary 4 is a second generalization of Glazer and Rubinstein's theorem.

4 Credibility

Glazer and Rubinstein (2006) establish that in the case of two actions, there is an optimal credible rule. This result is robust to some variations to the model. For example, Glazer and Rubinstein (2004) presented a credibility result for a related model in which the listener has two actions, and in which the speaker sends a cheap talk message to the listener, after which the listener can verify some but not all the information provided by the speaker. Likewise Glazer and Rubinstein (2001) presented an example of a two speaker debate such that the optimal debate from the listener's perspective was credible.

Glazer and Rubinstein (2006) point out that in the case of many actions, the credibility result no longer holds. The following example is adapted from their paper:

Example 2 Suppose that there are two types $T = \{t_1, t_2\}$, and the probability distribution π is such that $\pi(t_1) = .4$ and $\pi(t_2) = .6$. There are three actions, where again I write a_i instead of i. Assume, moreover that $u(a_i) = i$. The message correspondence is given by $M(t_1) = \{m_1\}, M(t_2) = \{m_1, m_2\}$. The listener's utility function is:

v	a_1	a_2	a_3
t_1	0	-1	1
t_2	0	1	-1

The rule f which assigns a_1 to m_1 and a_2 to m_2 is the unique deterministically optimal rule. However, in this case, upon seeing m_1, the listener would know that the speaker's type is t_1, and therefore would prefer to take action a_3. So f is not credible.[1] Allowing for random rules, the optimal rule f' responds to m_1 by randomizing over a_1 and a_3 with equal probability, and responds to m_2 by taking a_2 with probability 1. Again, this rule is not credible. If the speaker chooses the best reply which is optimal from the speaker's perspective, type t_1 will send m_1 and type t_2 will send m_2. But in this case the listener will prefer to select a_3 with probability 1 upon seeing m_1. So f' is not credible either. $\square$

Observe that in the above example, the listener's utility function is "convex" in actions given t_1. One can restore the credibility result under an appropriate concavity assumption. I also take the opportunity to show that optimality is actually a stronger assumption than is necessary for establishing the credibility result. One only needs to check that the persuasion rule is "locally optimal" in a sense to be defined below.

---

[1] Although, they do not say so explicitly, Glazer and Rubinstein's analysis of a counter-example shows that with many actions, a *deterministically* optimal persuasion rule may not be credible. This is natural, because they are making a comparison with the two action case, where deterministic optimality implies optimality. Here I extend the example to consider optimal rules, which in this case are not deterministic.

In what follows, I will discuss the credibility of deterministically optimal persuasion rules. Recall that deterministic optimality implies optimality in the case of two actions or as shown in Corollary 4, when the listener's utility function is a concave transformation of the speaker's utility function given any speaker type.

In order to define the notion of local optimality mentioned above, I will define a correspondence $N : D \to 2^D$. For any deterministic persuasion rule f, we may think of $N(f)$ as a discrete neighborhood of f. N is defined as follows:

$$N(f) := \left\{ \begin{array}{c} f' \in D : \exists \text{ adjacent } a, a' \in A, \forall m \in \mathbf{M}, \\ f'(m) \neq f(m) \Rightarrow f(m) = a \text{ and } f'(m) = a'. \end{array} \right\}$$

Thus, $f' \in N(f)$ if f' can be arrived at by taking some set of messages which f assigns to a and moving them to an adjacent action a'. On the other hand if a and a' are adjacent, and and f and f' differ by the interchange of two nonempty sets of messages K and K' which f assigns to a and a' respectively and f' assigns to a' and a respectively, then $f' \notin N(f)$. If f' differs from f by the movement of messages to nonadjacent actions, or by the movement of messages initially assigned to more than one action, then $f' \notin N(f)$. Likewise if messages reassigned to several actions.

Observe that for all $f, f' \in D$, (i) $f \in N(f)$ and (ii) $f' \in N(f) \Leftrightarrow f \in N(f')$. So the relation $f' \in N(f)$ is reflexive and symmetric, but it is not transitive.

Definition 2 *$f \in D$ is* ***locally deterministically optimal*** *if:*

$$f \in \text{argmax}_{f' \in N(f)} \max_{\sigma \in B(f')} \sum_{t \in T} \sum_{m \in \mathbf{M}} \sigma_{t,m} v(f'(m), t) p(t)$$

In the above definition "$\max_{\sigma \in B(f')}$" could be replaced with "$\min_{\sigma \in B(f')}$" or the choice of any σ in $B(f')$ because all speaker best replies to a deterministic persuasion rule result in the same utility for the listener. Local deterministic optimality does not imply deterministic optimality, much less optimality. This is true even in the case of two actions:

Example 3 Suppose that there are two actions with 3 types $\{t_1, t_2, t_3\}$. Suppose that $\mathbf{M} = \{m_1, m_2\}$, and $M(t_1) = \{m_1, m_2\}, M(t_2) = \{m_1\}, M(t_3) = \{m_2\}$. Suppose that each type occurs with equal probability. Suppose that the listener's utility function is given by the following table:

v	a_1	a_2
t_1	0	3
t_2	0	-2
t_3	0	-1

Then the persuasion rule f which assigns a_2 to m_1 and a_1 to m_2 is locally deterministically optimal even though it is worse than the rule f' which assigns a_1 to m_1 and a_2 to m_2. This is because in moving from f to f', one must lower the action assigned to one message, and raise the action assigned to another.

As a counter-example exists to the proposition that deterministically optimal rules can be credibly implemented with many actions, an additional assumption must be imposed to attain a credibility result with many actions.

Assumption 2 *There exists a strictly increasing function* $r : A \to \mathbb{R}$ *and for all* $t \in T$, *there exists a concave function* $c_t : \mathbb{R} \to \mathbb{R}$ *such that for all* $a \in A$, $v(a,t) = c_t(r(a))$.

Observe that Assumption 2 is always satisfied in the case of two actions. Observe that Assumption 2 is weaker than Assumption 1. With many actions, if r can be chosen to be equal to the speaker's utility function u (as in Assumption 1), then by Corollary 4, all optimally deterministic rules are optimal. It is interesting that an assumption which is sufficient for optimality of deterministic rules is similar to an assumption which is sufficient for credibility of rules which are optimal *among* deterministic rules.

Theorem 2 *Given Assumption 2, every locally deterministically optimal persuasion rule is credible.*

Proof. In appendix.

Observe that for any deterministic persuasion rule f^* and speaker strategy σ^*, in order to check that f^* is optimal *against* σ^*, it is sufficient to check that changing the response to any single message m would not improve the listener's utility, and given Assumption 2, it is sufficient to check that changing the response to one message to an adjacent action does not improve the listener's utility. Thus one only has to consider deviating to a proper subset of $N(f^*)$, although the calculation of the listener's payoff is done holding fixed a particular speaker strategy σ^*. This is true in particular when verifying that (σ^*, f^*) is an equilibrium. In contrast, verification that a deterministic persuasion rule is locally deterministically optimal amounts to checking all deviations in $N(f^*)$, however, not holding any speaker strategy σ fixed, but rather assuming that σ varies so as to remain a best reply to $f \in N(f^*)$. What Theorem 2 says is that this latter test is sufficient for the existence of a speaker strategy $\sigma^* \in B(f^*)$ such that f^* passes the former test with respect to σ^*.

Corollary 5 *Given Assumption 1, there exists an optimal persuasion rule which is also credible.*

Proof. This follows from theorem 2 and Corollary 4, and the fact that every deterministically optimal rule is also locally deterministically optimal. □

5 Conclusion

This paper studied optimal persuasion rules. Conditions were found under which optimal rules were deterministic and credible with many actions. Also the property of quasi-determinism was shown to characterize some optimal persuasion rule in every persuasion problem. It was shown that a discrete notion of local optimality was sufficient for credibility.

6 Appendix

6.1 Proof of Theorem 1

Proof. Suppose that $f \in R(\sigma)$ is such that for some $m^*, \sum_{a\in A} u(a) f_{m^*,a} \notin \{u(a) : a \in A\}$. (i.e., f violates (1)). Let

$$P := \{m \in \mathbf{M} : \sum_{a\in A} u(a) f_{m,a} = \sum_{a\in A} u(a) f_{m^*,a}\}$$

For each $m \in P$, let $\underline{a}^m$ (resp., $\overline{a}^m$) be the speaker's least (resp., most) preferred action a such that $f_{m,a} > 0$. The fact that every $m \in P$ gives the speaker an expected utility outside of $\{u(a) : a \in A\}$ implies that $\underline{a}^m \neq \overline{a}^m$. Now define two persuasion rules $\underline{f}$ and $\overline{f}$ such that for all $m \in P$:

$$\underline{f}_{m,\underline{a}^m} = f_{m,\underline{a}^m} + \epsilon^m, \quad \underline{f}_{m,\overline{a}^m} = f_{m,\overline{a}^m} - \epsilon^m, \quad \overline{f}_{m,\overline{a}^m} = f_{m,\overline{a}^m} + \epsilon^m, \quad \overline{f}_{m,\underline{a}^m} = f_{m,\underline{a}^m} - \epsilon^m \ ,$$

where for all $m \in P$, ϵ^m solves:

$$(u(\overline{a}^m) - u(\underline{a}^m))\epsilon^m = \epsilon$$

for some $\epsilon > 0$. In all other cases, $\overline{f}$ and $\underline{f}$ coincide with f. If ϵ is chosen small enough then (i) $\overline{f}$ and $\underline{f}$ are in fact persuasion rules (i.e., assign a probability distribution over actions to each message) and, (ii) the ranking of messages (including ties) according to the speaker's expected utility to sending them is the same under $f, \underline{f}$, and $\overline{f}$. It follows that $\underline{f}$ and $\overline{f}$ are in $R(\sigma)$. On the other hand $f = \frac{1}{2}\underline{f} + \frac{1}{2}\overline{f}$. So f is not an extreme point of $R(\sigma)$. So any extreme point of $R(\sigma)$ satisfies (1).

Next, assume that $f \in R(\sigma)$ violates (2). Then there exist m^* and distinct $a', a'', a''' \in A$ such that for all $a \in B := \{a', a'', a'''\}, f_{m^*,a} > 0$. Define $e := \sum_{a\in B} u(a) h_a$, where $h_a := f_{m,a}/(\sum_{b\in B} f_{m^*,b})$. Notice that since the numbers, $u(a'), u(a'')$, and $u(a''')$ are all distinct, the set $E := \{r \in \mathbb{R}^B : e = \sum_{a\in B} u(a) r_a, \sum_{a\in B} r_a = 1\}$ is a one-dimensional affine space. Moreover, there must be two distinct pairs of elements in the set $\{u(a'), u(a''), u(a''')\}$ such that e is a (possibly degenerate for one pair) weighted average of the pair. So assume wlog that there exist $\alpha, \beta \in [0,1)$ such that $e = \alpha u(a') + (1-\alpha)u(a'')$ and $e = \beta u(a') + (1-\beta)u(a''')$. In other words, $(\alpha, 1-\alpha, 0), (\beta, 0, 1-\beta) \in E$. Moreover $(\alpha, 1-\alpha, 0), (\beta, 0, 1-\beta)$ are affinely independent. So since $(h_{a'}, h_{a''}, h_{a'''}) \in E$ and $h_a > 0$ for all $a \in B$, there exists $\gamma \in \mathbb{R}$ such that $\gamma(\alpha, 1-\alpha, 0) + (1-\gamma)(\beta, 0, 1-\beta) = (h_{a'}, h_{a''}, h_{a'''})$. Since $(h_{a'}, h_{a''}h_{a'''}) \geq 0$, it follows that $\gamma \in (0,1)$. So define persuasion rules f' and f'' so that:

$$\begin{array}{lll} f'_{m^*,a'} = \alpha \sum_{a\in B} f_{m^*,a}, & f'_{m^*,a''} = (1-\alpha) \sum_{a\in B} f_{m^*,a}, & f'_{m^*,a'''} = 0 \\ f''_{m^*,a''} = \beta \sum_{a\in B} f_{m^*,a}, & f''_{m^*,a''} = 0, & f''_{m^*,a'''} = (1-\beta) \sum_{a\in B} f_{m^*,a} \end{array}$$

In all other cases, f' and f'' coincide with f. Notice that the speaker's expected utility to any message is the same under f' and f'' as under f. It then follows from the fact that $f \in R(\sigma)$ that both f' and f'' are in $R(\sigma)$. Moreover $f = \gamma f' + (1-\gamma) f''$. So f is not an extreme point of $R(\sigma)$. So any extreme point of $R(\sigma)$ satisfies (2). □

6.2 Proof of Theorem 2

Define:

$$T_a \quad := \quad \{t \in T : \max\{f(m) : m \in M(t)\} = a\}$$

For all $t \in T_a$, define:

$$M^*(t) \quad := \quad M(t) \cap f^{-1}(a)$$

Thus $M^*(t)$ can be thought of as a message correspondence which only allows speaker types to send messages which give that type the highest action given f.

For each $a \in A \setminus \{\underline{a}\}$, define $v_t^a := \frac{v(a,t) - v(a-1,t)}{r(a) - r(a-1)} p(t)$, and define $v_t^{\overline{a}+1} := -\infty, v_t^{\underline{a}} := \infty$, where ∞ is not literally infinity, but rather a very large number.

Choose $a \in A$. Then credibility implies that the following system of equations has a solution (σ^a, γ):

$$\begin{array}{rl} \forall t \in T_a, \forall m \in M^*(t), & \sigma^a_{t,m} \geq 0 \\ \forall m \in f^{-1}(a), & \gamma^+_m, \gamma^-_m \geq 0 \\ \forall t \in T_a, & \displaystyle\sum_{m \in M^*(t)} \sigma^a_{t,m} = 1 \\ \forall m \in f^{-1}(a), & \left\{ \begin{array}{l} \sum_{t \in T_a^m} v_t^a \sigma^a_{t,m} - \gamma^+_m = 0 \\ \\ \sum_{t \in T_a^m} (-v_t^{a+1}) \sigma^a_{t,m} - \gamma^-_m = 0 \end{array} \right. \end{array}$$

where $T_a^m := \{t \in T_a : m \in M^*(t)\}$. I will refer to the above system of equations as S_a. The Farkas alternative for S_a is:

$$\forall t \in T_a, \forall m \in M^*(t), \qquad x_m^+ v_t^a - x_m^- v_t^{a+1} + y_t \geq 0 \tag{3}$$

$$\sum_{t \in T_a} y_t < 0 \tag{4}$$

$$\forall m \in f^{-1}(a), \qquad x_m^-, x_m^+ \leq 0 \tag{5}$$

Now assume for contradiction that (3)-(5) has a solution (x, y). Let $x_m = x_m^+ - x_m^-$. It follows from Assumption 2, (3), and (5) that $\forall t \in T_a, \forall m \in M(t)$:

$$x_m v_t^a + y_t \geq 0 \tag{6}$$

$$x_m v_t^{a+1} + y_t \geq 0 \tag{7}$$

Define:

$$M^+ := \{m \in f^{-1}(a) : x_m > 0\} \quad , \quad M^- := \{m \in f^{-1}(a) : x_m \leq 0\}$$

For $N \subseteq \mathbf{M}$, define:

$$\begin{aligned} \tau(N) &:= \{t \in T : M^*(t) \subseteq N\} \\ \nu(N) &:= \{t \in T : M^*(t) \cap N \neq \emptyset\} \end{aligned}$$

Observe that local optimality implies that for all $a \in A$ and for all $N \subseteq f^{-1}(a)$:

$$\sum_{t \in \tau(N)} v_t^a \geq 0 \tag{8}$$

$$\sum_{t \in \nu(N)} v_t^{a+1} \leq 0 \tag{9}$$

(8) says that the listener would not have an incentive to reassign any set of messages which lead to a to $a-1$, and (9) says that he would not have an incentive to reassign them to $a+1$.

Define:

$$\Phi := \{\varphi \in [f^{-1}(a)]^{T_a} : \forall t \in T_a, \varphi(t) \in M^*(t)\}$$

Lemma 1 $\forall K^+ \subseteq M^+, \forall K^- \subseteq M^-$ s.t. $K^+, K^- \neq \emptyset, \exists \varphi \in \Phi, \exists \overline{m} \in M^+, \exists \underline{m} \in M^-$,

$$\sum_{t \in \nu(K^+)} x_{\underline{\varphi}(t)} v_t^{a+1} \leq x_{\overline{m}} \sum_{t \in \nu(K^+)} v_t^{a+1} \tag{10}$$

$$\sum_{t \in \tau(K^-)} x_{\overline{\varphi}(t)} v_t^a \leq x_{\underline{m}} \sum_{t \in \tau(K^-)} v_t^a \tag{11}$$

Proof. First suppose that $|K^+| = |K^-| = 1$, so that we may write $K^+ = \{\overline{m}\}, K^- = \{\underline{m}\}$. Notice that all $\varphi \in \Phi$ and all $t \in \tau(K^-), \varphi(t) = \underline{m}$. So (11) will always be satisfied. On the other hand, we may choose φ so that for all $t \in \nu(K^+), \varphi(t) = \overline{m}$, thus ensuring that (10) is satisfied.

We now proceed by induction. Assume that (10)-(11) are satisfied whenever $|K^+| \leq n^+$ and $|K^-| \leq n^-$. Then we will argue that these equations are satisfied whenever $|K^+| \leq n^+ + 1$ and $|K^-| \leq n^-$, and also whenever $|K^+| \leq n^+$ and $|K^-| \leq n^- + 1$.

So consider $K'^+ \subseteq M^+$, and let $\overline{m}'$ be such that:

$$x_{\overline{m}'} = \min\{x_m : m \in K'^+\} \tag{12}$$

Let $K^+ =: K'^+ \setminus \{\overline{m}'\}$. Now, appealing to the inductive hypothesis, consider $x_{\overline{m}}, \varphi$ satisfying (10)-(11)for K^+ (and any K^-). Choose $\varphi' \in \Phi$ such that φ' agrees with φ on $\nu(K^+) \cup \tau(K^-)$. Notice that for all $t \in \nu(K'^+) \setminus \nu(K^+)$, $\overline{m}' \in M^*(t)$. So for all such t choose $\varphi'(t) = \overline{m}'$. Notice that since $\nu(K'^+) \cap \tau(K^-) = \emptyset$, this does not interfere with the behavior of φ on $\tau(K^-)$, and so (11) is still satisfied. We have:

$$\begin{aligned}
\sum_{t \in \nu(K'^+)} x_{\varphi'(t)} v_t^{a+1} &= x_{\overline{m}'} \sum_{t \in \nu(K'^+) \setminus \nu(K^+)} v_t^{a+1} + \sum_{t \in \nu(K^+)} x_{\varphi(t)} v_t^{a+1} \\
&\leq x_{\overline{m}'} \sum_{t \in \nu(K'^+) \setminus \tau(K^+)} v_t^{a+1} + x_{\overline{m}} \sum_{t \in \nu(K^+)} v_t^{a+1} \\
&\leq x_{\overline{m}'} \sum_{t \in \nu(K'^+)} v_t^{a+1},
\end{aligned}$$

where the first inequality follows from the inductive hypothesis, and the second follows from (9) and (12).

A similar argument with the roles of τ and ν reversed, applies when we increase the size of K^- by one. Again the argument begins by defining $\underline{m}'$ so that $x_{\underline{m}'} = \max\{x_m : m \in K'^-\}$. The reason that max here plays the role of min in the previous argument is that the inequalities in (10) and (11) point in opposite directions. $\square$

Now apply the lemma to $K^+ = M^+$ and $K^- = M^-$, and acquire the corresponding objects $\varphi \in \Phi$, $\overline{m} \in M^+$, and $\underline{m} \in M^-$. Observe that $\tau(M^-)$ and $\nu(M^+)$ partition T_a. So from (4), either $\sum_{t \in \tau(M^-)} y_t < 0$ or $\sum_{t \in \nu(M^+)} y_t < 0$. Using (6) and (7), this implies that either:

$$\sum_{t \in \tau(M^-)} x_{\varphi(t)} v_t^a > 0, \tag{13}$$

or

$$\sum_{t \in \nu(M^+)} x_{\varphi(t)} v_t^{a+1} > 0. \tag{14}$$

However, noting that $x_{\underline{m}} \leq 0$, (13) contradicts (11) and (8). Likewise, noting that $x_{\overline{m}} \geq 0$, (14) contradicts (10) and (9). It follows that (3)-(5) does not have a solution, which by the Farkas lemma, implies that S_a does have a solution. So if the speaker uses a strategy σ defined such that $\sigma_{t,m} = \sigma^a_{t,m}$ for $t \in T_a$, it follows that the listener could never benefit by reassigning a message m with $f(m) = a$ to either $a+1$ or $a-1$. Assumption 2 then implies that the listener could never benefit by reassigning message m to any other action. $\square$

References

Glazer, J., and A. Rubinstein (2001): "Debates and Decisions: On a Rationale of Argumentation Rules," *Games and Economic Behavior*, 36, 158–173.

——— (2004): "On the Optimal Rules of Persuasion," *Econometrica*, 72, 1715–1736.

——— (2006): "A Study in the Pragmatics of Persuasion: A Game Theoretical Approach," *Theoretical Economics*, 1, 395–410.