Selecting Cheap-Talk Equilibria*

Ying Chen, Navin Kartik, and Joel Sobel†

September 30, 2007

Abstract

There are typically multiple equilibrium outcomes in the Crawford-Sobel (CS) model of strategic information transmission. This paper identifies a simple condition on equilibrium payoffs, called NITS, that selects among CS equilibria. Under a commonly used regularity condition, only the equilibrium with the maximal number of induced actions satisfies NITS. We discuss various justifications for NITS, including perturbed cheap-talk games with non-strategic players or costly lying. We also apply NITS to other models of cheap-talk, illustrating its potential beyond the CS framework.

Journal of Economic Literature Classification Numbers: C72, D81.

Keywords: cheap-talk; babbling; equilibrium selection; almost-cheap talk.

*This paper supersedes portions of Chen's (2006) "Perturbed Communication Games with Honest Senders and Naive Receivers" and Kartik's (2005) "Information Transmission with Almost-Cheap Talk." We are grateful to Vince Crawford for encouraging us to write this paper, and David Eil, Sidartha Gordon, Sjaak Hurkens, Melody Lo, John Morgan, various seminar audiences, and three anonymous referees for comments. We thank Steve Matthews for making available an old working paper of his. For advice and support related to this research, Chen and Kartik are indebted to their respective dissertation supervisors: David Pearce, Stephen Morris, and Dino Gerardi (Chen); and Doug Bernheim and Steve Tadelis (Kartik).

†Chen: Department of Economics, Arizona State University, email: yingchen@asu.edu; Kartik and Sobel: Department of Economics, University of California, San Diego, email: nkartik@ucsd.edu and jsobel@ucsd.edu respectively. For financial support, Kartik and Sobel thank the National Science Foundation, and Sobel also thanks the Guggenheim Foundation and the Secretaría de Estado de Universidades e Investigación del Ministerio de Educación y Ciencia (Spain). For hospitality and administrative support, Kartik is grateful to the Institute for Advanced Study at Princeton, and Sobel is grateful to the Departament d'Economia i d'Història Econòmica and Institut d'Anàlisi Econòmica of the Universitat Autònoma de Barcelona.

1 Introduction

In the standard model of cheap-talk communication, an informed Sender sends a message to an uninformed Receiver. The Receiver responds to the message by making a decision that is payoff relevant to both players. Talk is cheap because the payoffs of the players do not depend directly on the Sender's message. Every cheap-talk game has a degenerate, "babbling" equilibrium outcome in which the Sender's message contains no information, and, on the equilibrium path, the Receiver's response is equal to her ex-ante optimal choice.

Crawford and Sobel (1982) (hereafter CS) fully characterize the set of equilibrium outcomes in a one-dimensional model of cheap-talk with conflicts of interest. CS demonstrate that there is a finite upper bound, N^*, to the number of distinct actions that the Receiver takes with positive probability in equilibrium, and that for each $N = 1, \ldots, N^*$, there is an equilibrium in which the Receiver takes N actions. In addition, when a monotonicity condition holds, CS demonstrate that for all $N = 1, \ldots, N^*$, there is a unique equilibrium outcome in which the Receiver takes N distinct actions with positive probability, and the ex-ante expected payoff for both Sender and Receiver is strictly increasing in N. The equilibrium with N^* actions is often called the "most informative equilibrium"[1] and is typically the outcome selected for analysis in applications.

Ex-ante Pareto dominance is not a compelling equilibrium-selection criterion, especially because it is necessarily the case in the CS model that different Sender types have opposing preferences over equilibria.[2] There has been some interest in developing alternative selection arguments for this model. Evolutionary arguments (e.g. Blume et al., 1993) have only limited ability to select equilibria, essentially because there are conflicts of interest between different Sender types. Standard equilibrium refinements based on Kohlberg and Mertens's (1986) strategic stability—even those, like Banks and Sobel (1987) and Cho and Kreps (1987), that have been especially designed for signaling games—have no power to refine equilibria in cheap-talk games. Since communication is costless, one can support any equilibrium outcome with an equilibrium in which all messages are sent with positive probability, so arguments that limit the set of out-of-equilibrium beliefs have no power to refine. Farrell (1993) developed a variation of belief-based refinements that does select equilibria in cheap-talk games. Farrell assumes that there are always unused messages and that, if certain conditions are met, these messages must be interpreted in specific ways. His notion of neologism-proof equilibria does refine the set of equilibria in CS's games. Unfortunately, neologism-proof equilibria do not generally exist and no outcome satisfies the criterion in the leading (quadratic preferences, uniform prior) CS example when there are multiple equilibrium outcomes.[3]

This paper identifies a novel criterion to select equilibria in CS (and related) cheap-talk games. The criterion we pose is called NITS, for *No Incentive to Separate*, and is

[1] This terminology is misleading because adding actions typically does not lead to a refinement in the Receiver's information partition.

[2] If there are multiple equilibrium outcomes, one type always receives his most preferred action in the babbling equilibrium.

[3] The related proposal of announcement-proofness by Matthews et al. (1991) also eliminates all of the equilibria in CS's leading example when there are multiple equilibrium outcomes, whereas Rabin's (1990) concept of credible rationalizability eliminates none of them.

formally defined in Section 3. An equilibrium satisfies NITS if the Sender of the lowest type weakly prefers the equilibrium outcome to credibly revealing his type (if he somehow could). This has the effect of imposing a constraint on the payoff to the lowest type of Sender, in addition to the usual incentive constraints for equilibrium. We show that equilibria satisfying NITS always exist in the CS game, and the criterion is selective; in particular, if the CS monotonicity condition holds, only the most informative equilibrium outcome survives.

In Section 4, we discuss the appeal of NITS. After some intuitive arguments, we provide formal justifications by considering perturbed versions of the CS model. We study two perturbations: one where both the Sender and Receiver may be "non-strategic," in which case they tell the truth and believe the message is truthful respectively; the other where the Sender incurs a "lying cost" of misreporting his type. In each case, we show that only cheap-talk equilibria satisfying NITS can be limits of pure-strategy equilibria of the perturbed games satisfying an appropriate monotonicity restriction, as the perturbations vanish. If one believes that cheap-talk equilibria should be robust to such kinds of perturbations (and accepts the other conditions), then this provides a distinct argument from ex-ante Pareto dominance for selecting the most informative equilibrium in common applications.

Section 5 concludes by applying NITS to some other models of cheap-talk.

2 The Model

We follow the development of Crawford and Sobel (1982), but modify their notation. There are two players, a Sender (S) and a Receiver (R); only S has private information. The Sender's private information or type, t, is drawn from a differentiable probability distribution function, $F(\cdot)$, with density $f(\cdot)$, supported on $[0,1]$. S has a twice continuously differentiable von Neumann-Morgenstern utility function $U^S(a,t)$,[4] where $a \in \mathbb{R}$ is the action taken by R upon receiving S's signal. The Receiver's twice continuously differentiable von Neumann-Morgenstern utility function is denoted by $U^R(a,t)$. All aspects of the game except t are common knowledge.

We assume that, for each t and for $i = R, S$, denoting partial derivatives by subscripts in the usual way, $U^i_1(a,t) = 0$ for some a, and $U^i_{11}(\cdot) < 0$, so that U^i has a unique maximum in a for each t; and that $U^i_{12}(\cdot) > 0$. For each t and $i = R, S$, $a^i(t)$ denotes the unique solution to $\max_a U^i(a,t)$. Assume that $a^S(t) > a^R(t)$ for all t. For $0 \leq t' < t'' \leq 1$, let $\bar{a}(t',t'')$ be the unique solution to $\max_a \int_{t'}^{t''} U^R(a,t)dF(t)$. By convention, $\bar{a}(t,t) = a^R(t)$.

The game proceeds as follows. S observes his type, t, and then sends a message $m \in M$ to R, where M is any infinite set. R observes the message and then chooses an action, which determines players' payoffs. A pure-strategy (perfect Bayesian) equilibrium consists of a message strategy $\mu : [0,1] \to M$ for S, an action strategy $\alpha : M \to \mathbb{R}$ for R,

[4] In CS, $U^S(\cdot)$ also depends on a bias parameter which measures the differences in the preferences of R and S. We suppress this parameter because we are not primarily interested in how changing preferences influences equilibria.

and an updating rule $\beta(t \mid m)$ such that

$$\text{for each } t \in [0,1], \mu(t) \text{ solves } \max_m U^S(\alpha(m), t), \tag{1}$$

$$\text{for each } m \in M, \alpha(m) \text{ solves } \max_a \int_0^1 U^R(a,t)\beta(t \mid m)dt, \tag{2}$$

and $\beta(t \mid m)$ is derived from μ and F from Bayes's Rule whenever possible. The restriction to pure-strategy equilibria is without loss of generality for our purposes.[5] We say that an equilibrium with strategies (μ^*, α^*) induces action a if $\{t : \alpha^*(\mu^*(t)) = a\}$ has positive prior probability.

CS demonstrate that there exists a positive integer N^* such that for every integer N with $1 \leq N \leq N^*$, there exists at least one equilibrium in which the set of induced actions has cardinality N, and moreover, there is no equilibrium which induces more than N^* actions. An equilibrium can be characterized by a partition of the set of types, $t(N) = (t_0(N), \ldots, t_N(N))$ with $0 = t_0(N) < t_1(N) < \ldots < t_N(N) = 1$, and signals m_i, $i = 1, \ldots, N$, such that for all $i = 1, \ldots, N-1$

$$U^S(\bar{a}(t_i, t_{i+1}), t_i)) - U^S(\bar{a}(t_{i-1}, t_i), t_i)) = 0, \tag{3}$$

$$\mu(t) = m_i \text{ for } t \in (t_{i-1}, t_i], \tag{4}$$

and

$$\alpha(m_i) = \bar{a}(t_{i-1}, t_i). \tag{5}$$

Furthermore, all equilibrium outcomes can be described in this way.[6] In an equilibrium, adjacent types pool together and send a common message. Condition 3 states that Sender types on the boundary of a partition element are indifferent between pooling with types immediately below or immediately above. Condition 4 states that types in a common element of the partition send the same message. Condition 5 states that R best responds to the information in S's message.

CS make another assumption that permits them to strengthen this result. For $t_{i-1} \leq t_i \leq t_{i+1}$, let

$$V(t_{i-1}, t_i, t_{i+1}) \equiv U^S(\bar{a}(t_i, t_{i+1}), t_i) - U^S(\bar{a}(t_{i-1}, t_i), t_i).$$

A (forward) solution to (3) of length k is a sequence $\{t_0, \ldots, t_k\}$ such that $V(t_{i-1}, t_i, t_{i+1}) = 0$ for $0 < i < k$ and $t_0 < t_1$.

Definition 1. The Monotonicity (M) Condition is satisfied if for any two solutions to (3), $\hat{t}$ and $\tilde{t}$ with $\hat{t}_0 = \tilde{t}_0$ and $\hat{t}_1 > \tilde{t}_1$, then $\hat{t}_i > \tilde{t}_i$ for all $i \geq 2$.

[5] Our assumptions guarantee that R's best responses will be unique, so R will not randomize in equilibrium. The results of CS (specifically, their Theorem 1) demonstrate that S can be assumed to use a pure strategy, and moreover, only a finite number of messages are needed.

[6] One caveat is in order. There can be an equilibrium where type 0 reveals himself and is just indifferent between doing this and sending a signal that he in the adjacent step. We ignore this equilibrium, since the set of actions it induces is identical to those in another equilibrium where type 0 instead pools with the adjacent step. This is why equilibria can be characterized by a strictly increasing sequence that solves (3) and the boundary conditions.

Condition (M) is satisfied by the leading "uniform-quadratic" example in CS, which has been the focus of many applications. CS prove that if Condition (M) is satisfied, then there is exactly one equilibrium partition for each $N = 1, \ldots, N^*$, and the ex-ante equilibrium expected utility for both S and R is increasing in N.

3 The NITS Condition

We are now ready to define the condition that plays the central role in this paper.

Definition 2. An equilibrium (μ^*, α^*) satisfies the *No Incentive to Separate* (NITS) Condition if $U^S(\alpha^*(\mu^*(0)), 0) \geq U^S(a^R(0), 0)$.

NITS states that the lowest type of Sender prefers his equilibrium payoff to the payoff he would receive if the Receiver knew his type (and responded optimally). We postpone a discussion of NITS to the next section. In this section we show that the condition has the power to select between CS equilibria.

We present three results, ordered in decreasing level of generality. The first result shows that the equilibria with the maximum number of induced actions satisfy NITS. It also shows that if the babbling equilibrium survives NITS, then all equilibria do. The second result refines this insight under the assumption that there is exactly one equilibrium partition with N induced actions, for each N between 1 and N^*. Under this assumption, there exists an $\hat{N}$ such that the equilibria that satisfy NITS are precisely those with at least $\hat{N}$ actions induced. The final proposition makes the stronger assumption that Condition (M) is satisfied; in this case, only the (unique) equilibrium outcome with N^* induced actions survives NITS. Combined, the results demonstrate that imposing NITS is compatible with existence of equilibrium and that NITS selects equilibria that are commonly studied in applications.

Proposition 1. *If an N-step equilibrium fails to satisfy NITS, then there exists an $(N+1)$-step equilibrium. Moreover, if an equilibrium satisfies NITS, then so will any equilibrium with a shorter first segment.*

Consequently, every equilibrium with N^* induced actions satisfies NITS, and there is at least one equilibrium that satisfies NITS.

Proof. We first prove that if an equilibrium does not satisfy NITS, then there exists an equilibrium with more induced actions. Suppose that $\tilde{t} = (\tilde{t}_0, \ldots, \tilde{t}_N)$ is an equilibrium partition. We claim that if the equilibrium does not satisfy NITS, then for all $n = 1, \ldots, N$ there exists a solution to (3), t^n, that satisfies $t^n_0 = 0$, $t^n_n > \tilde{t}_{n-1}$, and $t^n_{n+1} = \tilde{t}_n$. The proposition follows from the claim applied to $n = N$. We prove the claim by induction on n.

Since the partition $\tilde{t}$ does not satisfy NITS, it follows that $V(0, 0, \tilde{t}_1) < 0$. On the other hand, $V(0, \tilde{t}_1, \tilde{t}_1) > 0$, because $\bar{a}(0, \tilde{t}_1) < a^R(\tilde{t}_1) < a^S(\tilde{t}_1)$. Continuity implies that there exists $x_1 \in (0, \tilde{t}_1)$ such that $V(0, x_1, \tilde{t}_1) = 0$. Setting $t^1_0 = 0, t^1_1 = x_1$ and $t^1_2 = \tilde{t}_1$ proves the claim for $n = 1$.

Suppose the claim holds for all positive integers up to some k, where $k < N$. We must show that it holds for $k+1$. Since $\tilde{t}$ is a solution to (3) and $k < N$, it follows that

$$V(\tilde{t}_{k-1}, \tilde{t}_k, \tilde{t}_{k+1}) = U^S(\bar{a}(\tilde{t}_k, \tilde{t}_{k+1}), \tilde{t}_k) - U^S(\bar{a}(\tilde{t}_{k-1}, \tilde{t}_k), \tilde{t}_k) = 0. \tag{6}$$

Using (6), $U^S_{11} < 0$, and

$$\bar{a}(\tilde{t}_{k-1}, \tilde{t}_k) < a^R(\tilde{t}_k) < a^S(\tilde{t}_k), \tag{7}$$

we see that $\bar{a}(\tilde{t}_k, \tilde{t}_{k+1}) > a^S(\tilde{t}_k)$. By the induction hypothesis, $t^k_k > \tilde{t}_{k-1}$ and $t^k_{k+1} = \tilde{t}_k$. Therefore $\bar{a}(t^k_k, t^k_{k+1}) \in (\bar{a}(\tilde{t}_{k-1}, t^k_{k+1}), a^R(t^k_{k+1}))$. It follows from (7) that

$$U^S(\bar{a}(\tilde{t}_{k-1}, \tilde{t}_k), \tilde{t}_k) < U^S(\bar{a}(t^k_k, t^k_{k+1}), t^k_{k+1}). \tag{8}$$

From (6) and (8) it follows that $V(t^k_k, \tilde{t}_k, \tilde{t}_{k+1}) < 0$. Since $V(t^k_k, \tilde{t}_k, t^k_{k+1}) > 0$, we conclude that there is a unique $x_{k+1} \in (t^k_{k+1}, \tilde{t}_{k+1})$ such that $V(t^k_k, t^k_{k+1}, x_{k+1}) = 0$. That is, it is possible to find a solution to (3) in which the $(k+1)^{\text{th}}$ step ends at t^k_{k+1} and the $(k+2)^{\text{nd}}$ step ends at $x_{k+1} < \tilde{t}_{k+1}$. By continuity, we can find a solution to (3) whose $(k+1)^{\text{th}}$ step ends at any $t \in (t^k_{k+1}, 1)$. For one such t the $(k+2)^{\text{nd}}$ step will end at $\tilde{t}_{k+1}$. This proves the claim.

To prove the second part of the proposition, suppose that an equilibrium with initial segment $[0, t_1]$ satisfies NITS. Consequently, Sender type 0 weakly prefers $\bar{a}(0, t_1)$ to $a^R(0)$. Since $U^S(\cdot)$ is single peaked, $\bar{a}(0, t)$ is increasing in t, and $\bar{a}(0, 0) = a^R(0)$, it follows that Sender type 0 will weakly prefer $\bar{a}(0, t)$ to $a^R(0)$ for all $t \in [0, t_1]$. ■

Proposition 2. *If there is only one equilibrium partition with N induced actions for any $N \in \{1, \ldots, N^*\}$, then there exists $\hat{N} \in \{1, \ldots, N^*\}$ such that an equilibrium with N actions satisfies NITS if and only if $N \geq \hat{N}$.*

This means that under the assumption, a set of low-step equilibria don't satisfy NITS, and the complementary set of high-step equilibria do.

Proof. Consider a family of solutions $t(x) = (t_0(x), t_1(x), \ldots, t_{K(x)}(x))$ to (3) satisfying $t_0(x) = 0$ and $t_1(x) = x$ and such that there exists no $t \in [t_{K(x)}, 1]$ such that $V(t_{K(x)-1}, t_{K(x)}, t) = 0$. It can be verified that $K(\cdot)$ has range $\{1, \ldots, N^*\}$, changes by at most one at any discontinuity, and if x is a discontinuity point of $K(\cdot)$, $t_{K(x)}(x) = 1$, so that x is the first segment boundary of a $(K(x))$-step equilibrium partition. Since $K(1) = 1$, it follows that if $K(t) = N$, then for each $N' \in \{1, \ldots, N\}$, there is at least one equilibrium of size N' with first segment boundary weakly larger than t. This implies that under the assumption of the Proposition, if t is a first segment boundary of an N-step equilibrium partition, no $t' > t$ can be the first segment boundary of a $(N+1)$-step equilibrium. Consequently, an $(N+1)$-step equilibrium has a shorter first segment than an N-step equilibrium. The desired conclusion follows from Proposition 1. ■

Proposition 3. *If Condition (M) is satisfied, only the unique equilibrium partition with the maximum number of induced actions satisfies NITS.*

Proof. We show that if the equilibrium partition with N steps satisfies NITS, then there is no equilibrium with $N+1$ steps. Suppose that $\tilde{t} = (\tilde{t}_0 = 0, \ldots, \tilde{t}_N = 1)$ is an equilibrium partition. It follows that $\tilde{t}$ satisfies (3). If the equilibrium satisfies NITS,

$V(0,0,\tilde{t}_1) = U^S(\bar{a}(0,\tilde{t}_1),0) - U^S(a^R(0),0) \geq 0$. This implies that a vector $\hat{t}$ that solves (3) with $\hat{t}_0 = \hat{t}_1 = 0$ must satisfy $\hat{t}_2 \geq \tilde{t}_1$, and by (M), $\hat{t}_n \geq \tilde{t}_{n-1}$ for all $n \geq 1$. Thus, $\hat{t}$ can have no more than $N+1$ steps. Using (M) again, any vector t solving (3) with $t_0 = 0 < t_1$ satisfies $t_n > \hat{t}_n$ for all $n > 0$. Consequently, no such vector t can satisfy $t_{N+1} = 1$, which is the desired result. ■

The above results demonstrate that while NITS is most powerful when Condition (M) is satisfied, the criterion may reduce the set of equilibrium outcomes even when (M) fails, although not necessarily to a singleton. We illustrate this point in a discrete version of the model. Suppose that preferences are $U^i(a,t) = -(a-t-b^i)^2$, where $b^R = 0$ and $b^S \in (0, 1/4)$, and the prior probability of types 0, 1/2, and 1 is 1/3 each. There are at least three equilibrium outcomes: a babbling outcome in which R's takes the action 1/2; a semi-pooling outcome in which $t = 0$ reveals itself and induces action 0, while the other types pool together and induce action 3/4; and a fully revealing outcome. Both the semi-pooling and fully revealing outcomes satisfy NITS, whereas the babbling outcome does not.[7]

4 Discussion and Justifications

The previous section demonstrated that imposing NITS selects amongst cheap-talk equilibria. A basic intuition for NITS comes from standard signaling models. Consider signaling models, like game-theoretic versions of the canonical Spence (1973) model, in which signals are costly, S's preferences are monotonic in R's actions, and R's action is monotonically increasing in S's type. It is natural to think of the lowest type ($t = 0$) of S as the worst type: no other type would want to be thought of as the lowest type and the lowest type would prefer to be thought of as any other type over itself. In this situation, interpreting an unsent message as coming from $t = 0$ is the least restrictive off-the-path belief. Any equilibrium outcome will remain an equilibrium outcome if one interprets unsent messages as coming from the lowest type. For this reason, NITS is always satisfied in Spencian models.

NITS would also hold in any equilibrium of a cheap-talk game in which S's preferences are strictly increasing in R's action, because in this case $a^R(0)$ is the type 0 Sender's least-preferred action. These games are not interesting, however, because the only equilibrium outcome is uninformative. Yet even when one assumes only that $a^S(t) > a^R(t)$ for all t (as in CS), S always prefers to reveal her type rather than be treated like the lowest type. Therefore, interpreting out-of-equilibrium messages as coming from the lowest type imposes what appears to be a weak restriction on the Receiver's behavior.

The first subsection below discusses the implications of NITS for out-of-equilibrium actions, and relates NITS to Farrell's (1993) notion of neologism-proofness. The second subsection points out that NITS will be satisfied if one endows the Sender with seemingly innocuous verifiable signals. The next two subsections study perturbations of the cheap-

[7]For completeness, we note that when $b^S < 1/8$, there is also a semi-pooling equilibrium in which $t = 1$ reveals itself and induces action 1, while the other types pool together and induce action 1/4. This equilibrium fails NITS. Moreover, because this is a discrete example, there are also equilibria in which one or more of the two lower Sender types induce multiple actions with positive probability.

talk game and show that NITS holds in these perturbed games within a class of equilibria, and that it is inherited in the limit as the perturbations vanish. The final subsection discusses how NITS may also be obtained by placing restrictions on feasible strategies, as in Lo (2006).

4.1 Novel Messages and Actions

Belief-based refinements of cheap-talk games require that the Sender can induce some novel actions by using novel messages. That is, these refinements assume that it is possible to induce actions that are not taken in equilibrium by using messages not sent in equilibrium. Every equilibrium outcome in CS that satisfies NITS can be supported by an equilibrium in which almost all types have a unique best response. In these equilibria, the Receiver responds to novel messages with an action in a (possibly degenerate) interval of $a^R(0)$. On the other hand, in an equilibrium that violates NITS, any rationalizable Receiver action that is not induced in equilibrium is strictly preferred by a positive measure of types to the action they induce in equilibrium. This implies that a CS outcome that violates NITS can only be generated by equilibria where the range of R's equilibrium strategy function is the set of induced actions. Hence, in non-NITS equilibria, novel messages cannot trigger novel actions. Intuitively, it is possible to have flexible interpretations of messages only when NITS holds.

Perhaps the best known belief-based refinement for cheap-talk games is Farrell's (1993) notion of neologism-proofness. Unfortunately, existence is not guaranteed by this criterion, which makes it problematic. Nevertheless, it is easily shown that CS equilibria which do not satisfy NITS are not neologism-proof.

Proposition 4. *An equilibrium that fails NITS is not neologism-proof.*

The proof in the Appendix shows that if an equilibrium does not satisfy NITS, there is some $t' > 0$ such that $[0, t']$ is a self-signaling set, i.e. if R interprets the message "my type is in $[0, t']$" literally and best responds to it, then it is precisely the types in $[0, t']$ that gain by sending the message relative to the equilibrium. It is on this basis that Farrell would reject an equilibrium that fails NITS. We note, however, that since Farrell rejects equilibria that contain *any* self-signaling sets, he also often rejects equilibria that do satisfy NITS, because one can also typically find self-signaling sets of the form $[t, 1]$.

4.2 Verifiable Information

Consider augmenting the CS game with a set of verifiable messages. In general, verifiable information could drastically affect the set of equilibrium outcomes. However, given that the basic incentive conflict in the game is that S desires to convince R that his type is higher than it actually is, it would seem innocuous to allow S to prove that his type is no greater than his true type. Specifically, suppose that in the message space M there exist subsets of messages M_t such that only S types with $t' \leq t$ can send messages in M_t. One can imagine that the sets M_t are strictly decreasing, so that there are simply more

things that low types can say.[8] In this environment, type $t = 0$ can reveal his identity by sending a message that is in M_0 but not in M_t for any $t > 0$. Type $t = 0$ would use such a message to destabilize any equilibrium that fails NITS. However, an equilibrium that survives NITS would survive the existence of this type of verifiable message. R could, for example, interpret any verifiable message as coming from $t = 0$.

Allowing "downward verifiability" therefore has the effect of imposing NITS, because it adds no new equilibrium outcomes and eliminates all equilibria that do not satisfy NITS. Given this effect, it is natural to ask what happens when, instead, the Sender can prove that his type is no lower than it really is. In this case, every cheap-talk equilibrium unravels. The $t = 1$ Sender would prove his type and improve his payoff (because $a^S(1) > a^R(1) > a$ for all actions a induced in equilibrium) and by induction, all types would reveal themselves in equilibrium. This is the standard intuition from disclosure models (e.g. Grossman, 1981; Milgrom, 1981).

The reason for the asymmetry is clear. When $a^S > a^R$ it is (locally) more desirable for S to try to inflate his type. Augmenting the message space with verifiable messages that are available only to higher types substantially changes the game because these extra messages make it impossible for S to exaggerate. On the other hand, given the direction of incentives to mimic, it is unexpected that S can gain from convincing R that his type is lower than it really is—but this is in fact the case in equilibria that violate NITS. Put differently, the fact that NITS has power to refine equilibria demonstrates that, in fact, sometimes S would like R to believe that his type is lower than she thinks it is in equilibrium. This happens when the lowest type of Sender is pooling with a large set of higher types.

4.3 Non-Strategic Players

Assume that the message space is $M = [0, 1]$, so that a message m may be interpreted as being a statement that the Sender's type is $t = m$. We introduce the possibility that players may be non-strategic: with independent probabilities, $\theta > 0$ and $\lambda > 0$ respectively, the Sender is an honest type who always tells the truth, i.e. sends $m = t$, and the Receiver is a non-strategic type who responds to a message as if it were truthful, i.e. plays $a = a^R(m)$. Otherwise the players act fully strategically and are called the "dishonest Sender" and the "strategic Receiver." This structure is common knowledge.

This perturbation induces a signaling game between the dishonest Sender and strategic Receiver, where messages are *not* cheap-talk. We refer the interested reader to Chen (2007b) for a detailed analysis of this model in the leading example of CS with uniform prior and quadratic preferences. She tackles existence and characterization of "message-monotone" equilibria—pure strategy (perfect Bayesian) equilibria where the dishonest

[8] It is not difficult to find applications that fit this framework. For example, a committee (the Sender) that deliberates behind closed doors may always want the public (the Receiver) to believe that it deliberated somewhat longer than it actually does. Since it can stay in its chambers for as long as it likes even after completing deliberations, it is impossible for it to prove that it actually did deliberate for the entire period up to the moment it announces its decision. Nevertheless, the announcement of its decision puts an upper bound on how long it deliberated.

Sender's strategy is weakly increasing in his type—for arbitrary $\theta > 0$ and $\lambda > 0$.[9] Our interest here concerns the limit of any such sequence of equilibria as the perturbations vanish.

Proposition 5. *As $\theta, \lambda \to 0$, the limit of any convergent sequence of message-monotone equilibria of the game with non-strategic players satisfies NITS.*

The proof is in the Appendix. It shows that NITS holds in any message-monotone equilibrium when $\theta > 0$ and $\lambda > 0$; consequently, the property holds in any limit of these equilibria. We note that some monotonicity restriction on equilibrium strategies in the perturbed game is necessary for this result.[10] In particular, consider a version of the CS game in which the type, message, and action spaces are all finite and preferences are quadratic loss functions. One can show that any equilibrium of the unperturbed game is the limit of equilibria (violating message monotonicity) in the perturbed game. More generally, any equilibrium outcome in this case is an element of a strategically stable component in the sense of Kohlberg and Mertens (1986).

4.4 Costly Lying

We now consider a direct payoff perturbation to CS, motivated by the notion that the Sender may have a preference for honesty. As before, assume the message space is $M = [0, 1]$. The game is identical to CS, except for Sender payoffs, which are given by $U^S(a, t) - kC(m, t)$, with $k > 0$. The function C is continuous, and for all t, $C(m, t)$ is strictly decreasing in its first argument at $m < t$ and strictly increasing in its first argument at $m > t$. This is naturally interpreted as a cost of lying for the Sender, minimized by telling the truth, e.g. $C(m, t) = |m - t|$. Plainly, when $k = 0$, the game is one of cheap-talk, but not so for any $k > 0$.

In such a model, with some more assumptions on the lying cost structure, Kartik (2007) studies the existence and characterization of "monotone equilibria," which are pure strategy (perfect Bayesian) equilibria where the Sender's strategy is weakly increasing in type—message monotonicity—and the Receiver's strategy is weakly increasing in the observed message—action monotonicity.[11] While he studies arbitrary $k > 0$, we are interested here in the limit as lying costs vanish.

Proposition 6. *As $k \to 0$, the limit of any convergent sequence of monotone equilibria of the game with costly lying satisfies NITS.*

[9] She shows that even when the dishonest Sender uses a weakly increasing strategy, the existence of the honest Sender creates non-monotonicities in the strategic Receiver's strategy.

[10] Conversely, monotonicity of strategies alone is not sufficient to make a selection; some perturbation is also needed. In particular, in CS, one can generate all equilibrium outcomes with equilibria where both Sender and Receiver use monotone strategies.

[11] Message monotonicity implies action monotonicity on the equilibrium path. Monotone equilibria also feature action monotonicity off the equilibrium path. Kartik (2005) shows that the ensuing Proposition only requires message monotonicity and a weak form of action monotonicity off the equilibrium path. His analysis also shows that under some assumptions on the lying cost function, action monotonicity implies message monotonicity in equilibrium.

The proof is in the Appendix. It shows that when k is sufficiently small, any monotone equilibrium must satisfy NITS; consequently, NITS is inherited in any limit of these equilibria. The restriction to monotone strategies plays an significant role in our proof of the Proposition. In general, we do not know whether NITS can be violated in the limit of some sequence of non-monotone equilibria as $k \to 0$.

4.5 Restrictions on Strategies

In deriving NITS from the above perturbations, monotonicity of strategies is an equilibrium refinement. In contrast, Lo (2006) imposes stronger monotonicity restrictions directly on strategies in a variation of the CS cheap-talk game and shows that these restrictions also yield NITS under iterated admissibility. The restrictions that Lo places are meant to describe limitations imposed by the use of a "natural language." She assumes that the message space is equal to the action space, which we assume are both $[a^R(0), a^R(1)]$.[12] This does not restrict the set of equilibrium outcomes but makes it possible to associate messages with their recommendations about Receiver actions, i.e. a message m is interpreted as a recommendation to take action $a = m$. Lo makes two restrictions on the strategies available to the Receiver. The first is that if a Receiver's strategy $\alpha(\cdot)$ ever induces the action a, then R interprets the message a literally (i.e., if there exists m such that $\alpha(m) = a$ then $\alpha(a) = a$). This restriction is considerably stronger than the monotonicity assumptions we use in the perturbations considered earlier, and indeed this property is not satisfied in the equilibria studied by Chen (2007b) and Kartik (2007). The second restriction is that the set of messages that induce a particular action is convex. These restrictions imply action monotonicity: $\alpha(\cdot)$ is weakly increasing. In turn, this implies that S also uses a message-monotone strategy: $\mu(\cdot)$ is weakly increasing. Lo's assumptions further guarantee an "absolute meaning" property: if two messages induce different actions, then the action induced by the higher (resp. lower) message is larger (resp. smaller) than the literal interpretation of the lower (resp. larger) message. Formally, if $m_1 < m_2$ and $\alpha(m_1) \neq \alpha(m_2)$, then $\alpha(m_2) > m_1$ and $\alpha(m_1) < m_2$.

In this framework, Lo demonstrates that under Condition (M), iterative deletion of weakly dominated strategies implies that any remaining strategy profile induces at least N^* actions from the Receiver (recall that N^* is the size of the maximal-step CS partition, which is the only CS outcome satisfying NITS under (M)).

The key observation is that in Lo's model, because of message monotonicity, weak dominance rules out any strategy for the Sender in which types close to 1 do not play the highest message $m = a^R(1)$.[13] Now suppose babbling does not satisfy NITS (in which case $N^* > 1$, by Proposition 1). Then because of action monotonicity and the absolute meaning property, sending message $\bar{a}(0,1)$ weakly dominates sending any $m > \bar{a}(0,1)$ for all types close to 0.[14] It is then iteratively dominated for the Receiver to respond

[12] Formally, Lo analyzes a discretized model, but this is not essential to the ensuing discussion.

[13] Similarly, types close to 1 send the highest available message in Chen (2007b) and Kartik (2007).

[14] This is true because absolute meaning says that if the Receiver plays a strategy that takes different actions for messages $m_1 = \bar{a}(0,1)$ and $m_2 > \bar{a}(0,1)$, then $\alpha(m_2) > \bar{a}(0,1)$. By action monotonicity, $\alpha(m_1) < \alpha(m_2)$. Since babbling violates NITS, type 0 and close enough types strictly prefer action

to message $\bar{a}(0,1)$ and $a^R(1)$ with the same action, which proves that any undeleted strategy for the Receiver must play at least 2 actions. Lo refines this argument to show that the Receiver must play at least N^* actions. Since her solution concept is iterated admissability rather than equilibrium, the Receiver may play even more than N^* actions.

5 Applications

We have formally defined the NITS condition only for the class of cheap-talk games satisfying the assumptions of CS. In order to extend the definition, we need a general notion of lowest type. In some applications, the identity of the lowest type is not clear.[15] Formulating NITS for such models is left to future research. Instead, in this section we describe two examples where the definition of the lowest type is clear and NITS can select equilibria as it does in the CS model.[16]

5.1 Veto Threats

Matthews (1989) develops a cheap-talk model of veto threats. This model frequently has two distinct equilibrium outcomes—one uninformative and one informative—and, we will show that under certain conditions, the natural adaptation of NITS selects the informative outcome.

In Matthews's model there are two players, a Chooser (C) and a Proposer (P). The players have preferences that are represented by single-peaked utility functions which we take to be of the form $-(a-b^i)^2$, where $a \in \mathbb{R}$ is the outcome of the game and $b^i \in \mathbb{R}$ is an ideal point for player $i = P, C$. The Proposer's ideal point $b^P = 0$ is common knowledge. The Chooser's ideal point is $b^C = t$, where t is his private information, drawn from a prior distribution that has a smooth positive density on a compact interval, $[\underline{t}, \bar{t}]$. The game form is simple: the Chooser learns his type, then sends a cheap-talk signal to the Proposer, who responds with a proposal, followed by which the Chooser either accepts or rejects the proposal. Accepted proposals become the outcome of the game. If the Chooser rejects the proposal, then the outcome is the status quo point $s = 1$.[17] When all Chooser types are at least one, the game is trivial (the status quo will be the final outcome). When all Chooser types prefer 0 to s, the game is trivial (the final outcome will be 0). We rule out these trivial cases by assuming that $\underline{t} < 1$ and $\bar{t} > \frac{1}{2}$. Matthews allows more general preferences and prior distributions. Only the uniqueness

$\alpha(m_1)$ to action $\alpha(m_2)$.

[15] For example, Gordon (2007) studies a variation of the CS model in which he permits $a^S(t) = a^R(t)$ for some t. This model allows the possibility of "inward bias" where $a^S(0) > a^R(0)$ and $a^S(1) < a^R(1)$, so that there are two candidates for the type that no other type wishes to imitate.

[16] The examples that follow have one-sided private information, as in CS. In a model with two-sided private information, Chen (2007a) shows that informative equilibria satisfy NITS and that the babbling equilibrium satisfies NITS only if it is the unique equilibrium outcome.

[17] In the final stage of the game, the Chooser decides to accept or reject a proposal under complete information. By replacing this decision by the optimal choice, one can reduce Matthews's model into a simple Sender-Receiver game, where the Chooser plays the role of Sender and the Proposer that of Receiver. This game does not satisfy the assumptions of Crawford and Sobel's (1982) model, however. In particular, the Proposer's preferences are not continuous in the Proposer's strategy.

result below depends on the quadratic specification of preferences. Matthews also uses a different normalization of b^P and s that has no substantive effect on the analysis.

As usual in cheap-talk games, this game has a babbling outcome in which the Chooser's message contains no information and the Proposer makes a single, take-it-or-leave-it offer that is accepted with probability strictly between 0 and 1. Matthews shows there may be equilibria in which two outcomes are induced with positive probability (size-two equilibria), but size $n > 2$ (perfect Bayesian) equilibria never exist. In a size-two equilibrium, P offers her ideal outcome to those types of C whose message indicates that their ideal point is low; this offer is always accepted in equilibrium. If C indicated that his ideal point is high, P makes a compromise offer that is sometimes accepted and sometimes rejected. Size-two equilibria only exist if $\underline{t}$ prefers $b^P = 0$ to $s = 1$, i.e. $\underline{t} < \frac{1}{2}$.

The NITS condition requires that one type of informed player do at least as well in equilibrium as it could if it could fully reveal its type. In CS, we imposed the condition on the lowest type, $t = 0$. It makes sense to apply the condition on the lowest type in Matthews's model as well. Intuitively, this is because higher types have more credible veto threats since the status quo is higher than the Proposer's ideal point and higher types like higher outcomes. Formally, let $a^P(t)$ be the action that the Proposer would take if the Chooser's type were known to be t.[18] In CS, when $t' > t$, Sender t' strictly prefers $a^R(t')$ to $a^R(t)$; when $t' < t$, Sender t' may or may not prefer $a^R(t)$ to $a^R(t')$, but there always exists a $t' < t$ with such preferences. In Matthews, when $t' > t$, Chooser t' weakly prefers $a^P(t')$ to $a^P(t)$. The preference is strict if $t' > 0$. When $t' < t$, Chooser t' may or may not prefer $a^P(t)$ to $a^P(t')$, but if $t' \in (0, 1)$, there always is such a type t. Hence in both models there is a natural ordering of types in which there is greater incentive to imitate higher types than lower types. In such an environment, there are fewer strategic reasons to prevent the lowest type from revealing itself, so the NITS condition is weakest when applied to the lowest type.

Consequently, we say that an equilibrium in Matthews's model satisfies NITS if the lowest type of Chooser, type $\underline{t}$, does at least as well as he would if he could reveal his type. Note that if the type-t Chooser reveals his type, then he will receive a payoff that is the maximum generated from the status-quo option, $s = 1$, and the Proposer's favorite outcome, $b^P = 0$. Thus, if a size-two equilibrium exists, it will satisfy NITS, because, as we observed earlier, in such an equilibrium type $\underline{t}$ will implement 0, whereas he can always implement 1. (This is an analog of our result that CS equilibria with the maximal number of actions will satisfy NITS.)

Furthermore, if a size-one equilibrium fails to satisfy NITS, then a size-two equilibrium must exist. (This is analogous to Proposition 1 for the CS model.) To see this, note that any Chooser can guarantee the status-quo outcome in equilibrium. Therefore, if a size-one equilibrium fails to satisfy NITS, the Chooser $\underline{t}$ must strictly prefer 0 to the offer made by the Proposer in the size-one equilibrium. Now for each t consider the preferences of a type-t Chooser who must select either 0 or the Proposer's optimal offer given that the Chooser's type is at least t. By assumption, when $t = \underline{t}$, this Chooser prefers 0. On the other hand, the proposal is preferable when $t = \bar{t}$. (If $\bar{t} < 1$, then this proposal is in $(0, \bar{t})$; if $\bar{t} \geq 1$, then this proposal can be taken to be 1.) Let $\tilde{a}(t')$ be a proposal that is

[18]The Proposer will offer her favorite outcome if the Chooser prefers this to the status quo, and something that leaves the Chooser indifferent to accepting the offer or the status quo otherwise.

optimal for the Proposer given that $t \in [t', \bar{t}]$. Continuity implies that there exists a $\tilde{t}$ such that $\tilde{t}$ is indifferent between 0 and the proposal given $t \in [\tilde{t}, \bar{t}]$. Hence there exists a size-two equilibrium in which types below $\tilde{t}$ send a message that induces the proposal 0.

Finally, under some conditions, the size-one equilibrium only satisfies NITS when no size-two equilibrium exists. (This is analogous to Proposition 3 for the CS model.) It is straightforward to check that a size-one equilibrium never satisfies NITS if $\underline{t} \leq 0$. If $\underline{t}$ prefers 1 to 0, then no size-two equilibrium exists. The interesting case is when $\underline{t} > 0$, prefers 0 to 1, and prefers the outcome in a size-one equilibrium to 0. A size-two equilibrium will fail to exist under these conditions if:

$$t \text{ prefers } a^P(t) \text{ to } 0 \text{ implies } t' > t \text{ prefers } a^P(t') \text{ to } 0. \tag{9}$$

This property need not hold without making further assumptions on preferences and the prior distribution. But, it appears to be a monotonicity condition similar to condition (M) from CS. While it is possible to derive a sufficient condition for (9), it is not especially instructive. Instead, we simply assert that it holds when preferences are quadratic and the prior is uniform.[19]

Finally, we note that neologism-proof outcomes (Farrell, 1993) often fail to exist in Matthews's model.[20] We omit the straightforward, but tedious, construction of the required self-signaling sets; the interested reader may refer to Matthews (1987). Lack of existence of neologism-proof equilibria in the model of veto threats parallels the lack of existence of neologism-proof outcomes in the CS model.

5.2 Signaling among Relatives

John Maynard Smith introduced the Sir Philip Sidney game to study signaling between related animals. The basic game is a two-player game with two-sided incomplete information and allows the possibility of costly communication. We describe how NITS can select a communicative outcome in a cheap-talk, one-sided incomplete information version of the model, based on Bergstrom and Lachmann (1998).

The Sender's type t is his fitness, which is private information to the Sender and drawn from a density $f(\cdot)$ supported on $[0, 1]$. After observing his type, the Sender sends a message m to the Receiver. The Receiver must then decide whether to transfer a resource to the Sender. If the Receiver transfers the resource, the Sender's direct benefit is 1 while the Receiver's direct benefit is $y \in (0, 1)$. If the Receiver does not transfer the resource, the Sender's direct benefit is t while the Receiver's direct benefit is 1. Total fitness is the weighted sum of a player's direct benefit and the benefit of the other, weighted by $k \in (0, 1]$.[21] Consequently, if a transfer is made with probability $1 - a$, then

[19] In this case $a^P(t) = \frac{2t-1+\sqrt{(2t-1)^2+3}}{3}$ since it is the solution to $\max -(\bar{t} - c) - a^2(c - t)$ subject to $c - a = 1 - c$. One can check that $a^P(t) > t$. Hence $a^P(t)$ is preferred to 0 if $2t > a^P(t)$ and condition (9) holds because $2t - a^P(t)$ is increasing.

[20] For Matthews's model, we say that an equilibrium outcome is neologism proof if there is no set of types T with the property that T is the set of types that strictly prefer the Proposer's optimal proposal when she knows that the Chooser's type lies in T to the equilibrium payoffs.

[21] In the biological context, k is the degree to which the players are related. In an economic context, k could be viewed as an altruism parameter.

$U^S(a,t) = (1-a)(1+ky) + a(t+k)$ while $U^R(a,t) = a(1+kt) + (1-a)(y+k)$. All aspects of the model except t are common knowledge.

This model does not satisfy the strict concavity assumption of CS, but otherwise is analogous, and it shares the property that optimal complete-information actions are (weakly) increasing in t. Provided that $y + k > 1$, which we assume to avoid triviality, both players benefit from (resp. are hurt by) transfers when t is low (resp. high), but the Sender prefers transfers for more values of t than the Receiver. Hence, in contrast to CS, the Sender likes weakly lower values of a than the Receiver for all t; accordingly, it is appropriate to apply NITS at $t = 1$. Since R prefers not to have a transfer of the resource when $t = 1$, an equilibrium satisfies NITS if and only if it induces $a = 1$ when $t = 1$.

By the linearity of preferences, there can be at most two actions induced in equilibrium. Define

$$y^* := \frac{y}{k} + 1 - \frac{1}{k}. \tag{10}$$

The Receiver finds it uniquely optimal to set $a = 0$ if $\mathbb{E}[t|m] < y^*$, uniquely optimal to set $a = 1$ if $\mathbb{E}[t|m] > y^*$, and is indifferent over all a otherwise.

As usual, a babbling equilibrium always exists. The babbling equilibrium satisfies NITS if and only if $\mathbb{E}[t] \geq y^*$. If an equilibrium with two induced actions exists, there must be a cutoff type, $t_1 \in (0,1)$, such that t_1 is indifferent between receiving or not receiving the transfer, which defines $t_1 = 1-k(1-y)$. Further, optimality of the Receiver's play requires that $\mathbb{E}[t|t < t_1] \leq y^*$ and $\mathbb{E}[t|t > t_1] \geq y^*$. The latter inequality necessarily holds, since by simple algebra, $t_1 \geq y^*$. Hence a two-step equilibrium exists if and only if

$$\mathbb{E}[t|t < 1 - k(1-y)] \leq y^*. \tag{11}$$

By the optimality of Receiver's play, if a two-step equilibrium exists, it satisfies NITS. Plainly, if the one-step equilibrium fails NITS, then a two-step equilibrium exists. If the one-step equilibrium satisfies NITS, a two-step equilibrium may or may not exist, depending on the prior density $f(\cdot)$. This conclusion is analogous to Proposition 1.[22]

[22]Unlike with Matthews's (1989) model, we do not have an analog here of Proposition 3.

A Appendix

Proof of Proposition 4. We need to show that an equilibrium failing NITS has a self-signaling set. Suppose that NITS fails at the equilibrium (μ^*, α^*) whose partition has first segment $[0, t_1]$. Then $U^S(a^R(0), 0) > U^S(\alpha^*(\mu^*(0)), 0)$, which implies that $\alpha^*(\mu^*(0)) > a^S(0)$. Since $\alpha^*(\mu^*(0)) < a^S(t_1)$, continuity implies that there is a $\tilde{t} \in (0, t_1)$ such that $a^S(\tilde{t}) = \alpha^*(\mu^*(0))$. Thus, $U^S(\bar{a}(0, \tilde{t}), \tilde{t}) < U^S(\alpha^*(\mu^*(0)), \tilde{t})$, and by continuity, there exists $t' \in (0, \tilde{t})$ such that

$$U^S(\bar{a}(0, t'), t') = U^S(\alpha^*(\mu^*(0)), t'). \tag{12}$$

Since $t' < \tilde{t}$ and $a^S(\tilde{t}) = \alpha^*(\mu^*(0))$, $U^S_{12} > 0$ and (12) imply that

$$U^S(\bar{a}(0, t'), t) > U^S(\alpha^*(\mu^*(0)), t) \text{ for all } t \in [0, t'); \tag{13}$$

$$U^S(\bar{a}(0, t'), t) < U^S(\alpha^*(\mu^*(0)), t) \text{ for all } t \in (t', 1]. \tag{14}$$

It follows from (13) and (14) that $[0, t']$ is a self-signaling set. ■

Proof of Proposition 5. Let $\mu^h(\cdot)$ denote the honest Sender's strategy and $\alpha^n(\cdot)$ denote the non-strategic Receiver's strategy. Then $\mu^h(t) = t$ and $\alpha^n(m) = a^R(m)$. Let $\mu(\cdot)$ and $\alpha(\cdot)$ denote the strategic players' strategies. The dishonest Sender's payoff if he sends m and induces action a from the strategic Receiver is $U^S_d(a, m, t) = \lambda U^S\left(a^R(m), t\right) + (1 - \lambda) U^S(a, t)$. It suffices to show that NITS holds in a message-monotone equilibrium for any $\theta, \lambda > 0$. Standard convergence arguments then imply that NITS is inherited in any limit as $\theta, \lambda \to 0$.

Fix arbitrary $\theta, \lambda > 0$. Suppose, to contradiction, that NITS does not hold in an message-monotone equilibrium, where the dishonest Sender and strategic Receiver play (μ^*, α^*). It follows from message monotonicity that the type 0 dishonest Sender must be pooling with higher types on message 0, since otherwise it could strictly benefit from deviating to sending message 0 (which would induce action $a^R(0)$ from either kind of Receiver). Let $t_1 = \sup\{t : \mu^*(t) = 0\} > 0$. Then, by message monotonicity, $\mu^*(t) = 0$ for all $t < t_1$, whereas $\mu^*(t) > 0$ for all $t > t_1$. Failure of NITS implies that $U^S_d(\alpha^*(0), 0, 0) < U^S_d\left(a^R(0), 0, 0\right)$ where $\alpha^*(0) = \bar{a}(0, t_1) > a^R(0)$. Since $U^S_{11} < 0$ and $a^S(0) > a^R(0)$, we have $\alpha^*(0) > a^S(0)$.

We claim that $\mu^*(\cdot)$ must be continuous at t_1. Suppose not. Then $\lim_{t \to t_1^+} \mu^*(t) > 0$. There exists an $\varepsilon > 0$ such that only the honest Sender sends ε and $\alpha^*(\varepsilon) = a^R(\varepsilon) \in (a^R(0), a^S(0))$. It follows that $U^S_d(\alpha^*(\varepsilon), \varepsilon, 0) > U^S_d\left(a^R(0), 0, 0\right) > U^S_d(\alpha^*(0), 0, 0)$ and the type 0 dishonest Sender strictly benefits from sending ε, a contradiction. A similar argument also establishes that $t_1 < 1$.

The continuity of $\mu^*(\cdot)$ at t_1 implies that there exists $t_2 > t_1$ such that $\mu^*(\cdot)$ is strictly increasing and continuous on (t_1, t_2) and $a^R(\mu^*(t)) < a^S(t)$ for all $t \in (t_1, t_2)$. Since $\alpha^*(\mu^*(t))$ is a weighted average of $a^R(\mu^*(t))$ and $a^R(t)$ for $t \in (t_1, t_2)$, $\alpha^*(\mu^*(t)) < a^S(t)$ as well. Moreover, $\alpha^*(\cdot)$ must be continuous and decreasing on $(\mu^*(t_1), \mu^*(t_2))$: continuous because because both μ^* and μ^h are continuous on the relevant domain, and decreasing to offset type t_1's incentive to deceive the non-strategic Receiver by sending some small message $\varepsilon > 0$. It follows that there exists an $m \in (0, \mu(t_2))$ such that $a^R(0) < a^R(m) < a^S(0)$ and $a^R(0) < \alpha^*(m) < \alpha^*(0)$. Therefore $U^S_d(\alpha^*(m), m, 0) >$

$U_d^S(\alpha^*(0), 0, 0)$ and the type 0 dishonest Sender strictly benefits from sending m, a contradiction. ■

Proof of Proposition 6. We will prove that when k is sufficiently small, any monotone equilibrium must satisfy NITS. Standard convergence arguments then imply that NITS must hold in the limit of any convergent sequence of monotone equilibria as $k \to 0$.

A basic implication of message monotonicity is that the set of types sending any given message is convex. Let t^* denote the type such that $U^S(\bar{a}(0, t^*), 0) = U^S(a^R(0), 0)$ if it exists, or $t^* = 1$ otherwise. It is straightforward that NITS is satisfied if and only if the highest type pooling with type 0 is no greater than t^*. Moreover, because of action monotonicity, NITS can only be violated if the lowest pool of types uses message 0. To see this, observe that if a monotone equilibrium violates NITS but does not have a pool of types using message 0, then by deviating to message 0, type 0 will elicit a weakly preferred response from the Receiver and will strictly save on lying cost, contradicting equilibrium.

Therefore, it suffices to show that the highest type using message 0 is no greater than t^* (if any), for small k. This is trivially true if $t^* = 1$, so assume henceforth $t^* < 1$. Suppose, towards contradiction, that for arbitrarily small k, there is a monotone equilibrium where the supremum of types pooling on message 0 is some $t_1^k > t^*$. We must have $t_1^k < 1$, for otherwise type 1 can profitably deviate up to message 1, because by action monotonicity, it will elicit a weakly higher response and strictly save on cost of lying. Note also that by considering k small enough, the difference in cost between sending any two messages for any type can be made arbitrarily small. Thus, for t_1^k to be indifferent between pooling on message 0 and mimicking a slightly higher type, there must be an interval of types, (t_1^k, t_2^k), that are pooling on some message $m_2 > 0$, with $U^S(\bar{a}(0, t_1^k), t_1^k) \approx U^S(\bar{a}(t_1^k, t_2^k), t_1^k)$.[23] Just as in CS, this requires that $\bar{a}(0, t_1^k) < a^S(t_1^k) < \bar{a}(t_1^k, t_2^k)$. But now, since by message monotonicity there are unused messages in $(0, m_2)$, and by action monotonicity these messages elicit actions that t_1^k weakly prefers to both $\bar{a}(0, t_1^k)$ and $\bar{a}(t_1^k, t_2^k)$, we must have $m_2 \leq t_1^k$: if not, type t_1^k can deviate to one of the unused messages, strictly save on message cost, and elicit a weakly preferred action. Note also that there is a positive lower bound on how small $t_2^k - t_1^k$ can be, by the CS assumptions on U^S and U^R.

If $t_2^k < 1$, then repeating the above logic inductively, we conclude that there must a finite N such that the N^{th} pool of types, $\langle t_{N-1}^k, 1]$, uses message $m_N \leq t_{N-1}^k < 1$. But then, type 1 can profitably deviate to message 1, eliciting a weakly preferred action (by action monotonicity), and saving on lying cost: contradiction with equilibrium. ■

[23] Type t_1^k must be pooling on message 0 or message m_2, since it cannot be separating because a type $t_1^k - \varepsilon$ would strictly prefer to mimic it.

References

Banks, J. S. and J. Sobel (1987): "Equilibrium Selection in Signaling Games," *Econometrica*, 55, 647–661.

Bergstrom, C. T. and M. Lachmann (1998): "Signaling among Relatives. III. Talk is Cheap," *Proceedings of the National Academy of Sciences, USA*, 95, 5100–5105.

Blume, A., Y.-G. Kim, and J. Sobel (1993): "Evolutionary Stability in Games of Communication," *Games and Economic Behavior*, 5, 547–575.

Chen, Y. (2007a): "Partially-informed Decision Makers in Games of Communication," Mimeo, Arizona State University.

——— (2007b): "Perturbed Communication Games with Honest Senders and Naive Recievers," In preparation, Arizona State University.

Cho, I.-K. and D. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179–221.

Crawford, V. and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50, 1431–1451.

Farrell, J. (1993): "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior*, 5, 514–531.

Gordon, S. (2007): "Informative Cheap Talk Equilibria as Fixed Points," Mimeo, Université de Montréal.

Grossman, S. J. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law & Economics*, 24, 461–483.

Kartik, N. (2005): "Information Transmission with Almost-Cheap Talk," Mimeo, University of California, San Diego.

——— (2007): "Strategic Communication with Costly Lying," In preparation, University of California, San Diego.

Kohlberg, E. and J.-F. Mertens (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1003–1037.

Lo, P.-y. (2006): "Common Knowledge of Language and Iterative Admissibility in a Sender-Receiver Game," Mimeo, Brown University.

Matthews, S. A. (1987): "Veto Threats: Rhetoric in a Bargaining Game," University of Pennsylvania, CARESS Working Paper # 87-06.

——— (1989): "Veto Threats: Rhetoric in a Bargaining Game," *Quarterly Journal of Economics*, 104, 347–369.

Matthews, S. A., M. Okuno-Fujiwara, and A. Postlewaite (1991): "Refining Cheap-Talk Equilibria," *Journal of Economic Theory*, 55, 247–273.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12, 380–391.

Rabin, M. (1990): "Communication between Rational Agents," *Journal of Economic Theory*, 51, 144–170.

Spence, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374.