On the Robustness of Informative Cheap Talk*

Ying Chen, Navin Kartik, and Joel Sobel†

March 23, 2007

Abstract

There are typically multiple equilibrium outcomes in the Crawford-Sobel (CS) model of strategic information transmission. This note 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 recent work on perturbed cheap-talk games. We also apply NITS to two 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.

*We thank Eddie Dekel for encouraging us to write this paper, Melody Lo for comments, and Steve Matthews for making available an old working paper.

†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. Sobel thanks the Guggenheim Foundation, NSF, and the Secretaría de Estado de Universidades e Investigación del Ministerio de Educación y Ciencia (Spain) for financial support. Sobel is grateful to the Departament d'Economia i d'Història Econòmica and Institut d'Anàlisi Econòmica of the Universitat Autònoma de Barcelona for hospitality and administrative support.

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 his 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. Because 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.[3]

Independent work by Chen (2006) and Kartik (2005) identifies a novel way to select equilibria in CS games. Both authors consider perturbed versions of the CS model,

[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 her 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, whereas Rabin's (1990) concept of credible rationalizability eliminates none of them.

show that equilibria exist in the perturbed games, and provide conditions under which only the CS equilibrium with N^* actions is the limit of equilibria to perturbed games as perturbations go to zero. If one believes that cheap-talk equilibria should be robust to the kinds of perturbations introduced in these papers (and accepts the other conditions), then these papers provide another argument, distinct from ex-ante Pareto dominance, for selecting the equilibrium with the most actions taken.

The same property plays an important role in the arguments of both Chen and Kartik. Provided that the CS monotonicity condition holds, if the Sender with the lowest type can credibly signal her type, then the most informative CS equilibrium (and only the most informative CS equilibrium) survives. The purpose of this note is to focus attention on this property, point out simple environments in which it will hold, and suggest why it may hold in other environments (such as those studied by Chen and Kartik).

Section 2 describes the model. Section 3 introduces the condition on equilibrium payoffs that generates the selection. Section 4 discusses environments in which the condition holds. Section 5 applies our condition to 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 (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, and an updating rule $\beta(t \mid m)$ such that

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

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

[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 $\beta(t \mid m)$ is derived from μ and F from Bayes's Rule whenever possible. Without loss of generality, we limit attention to pure-strategy equilibria.[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 $i \geq 2$.

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.

[5] Our assumptions guarantee that R's best responses will be unique, so R will not randomize in equilibrium. The results of CS 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.

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 her equilibrium payoff to the payoff she would receive if the Receiver knew her type (and responded optimally). Kartik (2005) introduced and named this condition. 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 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.

If the equilibrium 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 $\bar{a}(0, \tilde{t}_1) < a^R(\tilde{t}_1) < a^S(\tilde{t})$. 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 $1 \leq k < N$. We must show that it holds for $n = k+1$. 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 $t_{k+1}^k = \tilde{t}_k$. Therefore $\bar{a}(t_k^k, t_{k+1}^k) \in (\bar{a}(\tilde{t}_{k-1}, t_{k+1}^k), a^R(t_{k+1}^k))$. It follows from (7) that

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

From (6) and (8) we conclude that there is a unique $x_{k+1} \in (t_{k+1}^k, \tilde{t}_{k+1})$ such that $V(t_k^k, t_{k+1}^k, x_{k+1}) = 0$. That is, it is possible to find a solution to (3) in which the $(k+1)^{\text{th}}$ step ends at t_{k+1}^k 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 $t \in (t_{k+1}^k, 1)$. 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 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 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 a cheap-talk game satisfies (M), then only the equilibrium partition with the maximum number of induced actions satisfies NITS.*

Proof. We show that if the equilibrium 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. ■

4 Justifications for NITS

The previous section demonstrated that imposing NITS selects equilibria. In this section we discuss environments under which NITS will be satisfied.

A basic intuition comes from standard signaling models. Consider signaling models, like the canonical Spence 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.

Cheap-talk games are trivial when S's preferences over R's actions are independent of type. In this setting, all types of S would induce the same action (the one that they all prefer most) and therefore only the babbling equilibrium exists. The equilibrium satisfies NITS. Rather than assuming that S's preferences are monotonic in R's actions, CS only require that that $a^S(t) > a^R(t)$ for all t, so that all Sender types would like to be perceived as being higher than they truly are, but not necessarily desire to be perceived as the highest type. Hence no type would want to be mistaken for the $t = 0$ type over itself, although a type may be willing to send the same message as the lowest type even when other messages are available.

The first two subsections point out that NITS will be satisfied if one imposes seemingly innocuous restrictions on out-of-equilibrium beliefs or gives S weak ability to send verifiable signals. The final subsection summarizes the more sophisticated approaches of Chen (2006) and Kartik (2005). These papers view the CS cheap-talk game as the limit of games that admit the existence of equilibria in which at least the lowest types of S separate, 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.[7]

4.1 Restriction on Belief

A direct way to obtain NITS is to impose it as a restriction on beliefs, for example by assuming that there exists a message m^* such that upon observing m^*, the Receiver's beliefs are supported within $[0, t^*]$ where $t^* = 1$ if $U^S(\bar{a}(0,t),0) > U^S(a^R(0),0)$ for all $t \in (0,1)$ and t^* is the unique positive solution to $U^S(\bar{a}(0,t^*),0) = U^S(a^R(0),0)$ otherwise. This restriction on beliefs implies that NITS must hold in any equilibrium because the type 0 Sender will be able to guarantee a utility of at least $U^S(a^R(0),0)$ by sending the message m^*. (Note that when $t^* = 1$ the restriction holds trivially. In this case, there is no equilibrium that induces multiple actions.)

This restriction is in the spirit of Farrell's (1993) refinement. Like Farrell, it assumes that there are unused messages that are assigned a conventional meaning as long as that

[7] Goltsman and Pavolov (2006) characterize optimal communication mechanisms in the presence of an impartial mediator for the uniform-quadratic version of the CS model. They find that NITS holds in any optimal mechanism.

meaning is consistent with the incentives of the players. 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, $U^S(\bar{a}(0, \tilde{t}), 0) < U^S(\alpha^*(\mu^*(0)), 0)$, and by continuity, there exists $t' \in (0, \tilde{t})$ such that

$$U^S(\bar{a}(0, t'), t') = U^S(\alpha^*(\mu^*(0)), t'). \tag{9}$$

Since $t < \tilde{t}$ and $a^S(\tilde{t}) = \alpha^*(\mu^*(0))$, $U^S_{12} > 0$ and (9) imply that

$$U^S(\bar{a}(0, t'), t) > U^S(\alpha^*(\mu^*(0)), t) \text{ for all } t \in [0, t'); \tag{10}$$

$$U^S(\bar{a}(0, t'), t) < U^S(\alpha^*(\mu^*(0)), t) \text{ for all } t \in (t', 1]. \tag{11}$$

It follows from (10) and (11) that $[0, t']$ is a self-signaling set: 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. On this basis, Farrell would reject an equilibrium that fails NITS. But, since Farrell rejects equilibria that contain *any* self-signaling sets, he also rejects equilibria that do satisfy NITS. The belief restriction we need is weaker, and hence compatible with existence.

4.2 Verifiable Information

Augment the CS game with a set of verifiable messages. We consider two possibilities. In the first, each type of S can prove that her type is no greater than her true type. That is, in the message space M there exists subsets of messages M_t such that only S types with $t' \leq t$ can send message 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. In this environment, type $t = 0$ can reveal her 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. It adds no new equilibrium outcomes and eliminates all but the one that induces the most actions (when Condition (M) holds).

Given the effect of imposing "downward verifiability," it is natural to ask what happens when, instead, the Sender can prove that her type is no lower than it really is. In this case, the cheap-talk equilibrium unravels. The $t = 1$ Sender would prove her type and improve her 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 her 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 her 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 her type is lower than he thinks it is in equilibrium. This happens when the lowest type of Sender is pooling with a large set of higher types.

4.3 Almost-Cheap Talk

Chen (2006) and Kartik (2005) study distinct models of perturbed cheap-talk games: both assume literal messages spaces, but Chen augments CS by assuming that players may be non-strategic, whereas Kartik assumes that the Sender faces a cost of misreporting his type. The authors demonstrate the existence of a class of equilibria of their models and show that the limit of these equilibria as the perturbations vanish converge to CS equilibria satisfying NITS. Our goal here is to sketch the logic behind their arguments, elucidating why they lead to NITS, but deliberately finessing the intricate details of existence and convergence. The common idea is that in both models, limiting equilibria are such that the lowest message would entail a profitable deviations for the lowest type unless NITS is satisfied. Below, we unify this idea from both models by deriving contradictions in each case to the hypothesis that in the limit NITS is violated and yet there is no profitable deviation to the lowest message.

4.3.1 Non-Strategic Players

Chen (2006) studies the CS cheap-talk game with the assumption that $M = [0, 1]$ and augmented as follows: with small and independent probabilities, the Sender is an honest type (probability θ) who always tells the truth, i.e. sends $m = t$, and the Receiver is a naive type (probability λ) who always follows the messages as if they were truthful, i.e. plays $a = a^R(m)$. Otherwise the players act fully strategically and they are called the dishonest Sender and the strategic Receiver.

The limit of the perturbed games (as the probabilities of the non-strategic types go to zero) is the CS game. Chen looks at "message-monotone equilibria," i.e., pure strategy (perfect Bayesian) equilibria in which the dishonest Sender's strategy is weakly increasing in his type (state). She finds that NITS always holds in a message-monotone equilibrium if the probabilities of the non-strategic types are positive.

Let $\mu^h(\cdot)$ denote the honest Sender's strategy and $\alpha^n(\cdot)$ denote the naive 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 $U_d^S(a, m, t) = \lambda U^S\left(a^R(m), t\right) + (1 - \lambda) U^S(a, t)$.

We show by contradiction that NITS holds in a message-monotone equilibrium. Suppose not in an equilibrium (μ^*, α^*). Then 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 $U_d^S(\alpha^*(0), 0, 0) < U_d^S\left(a^R(0), 0, 0\right)$ where $\alpha^*(0) = \bar{a}(0, t_1) > a^R(0)$. Since $U_{11}^S < 0$ 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 $\mu^*(\varepsilon) = a^R(\varepsilon) \in (a^R(0), a^S(0))$. It follows that $U_d^S(\mu^*(\varepsilon), \varepsilon, 0) > U_d^S\left(a^R(0), 0, 0\right) > U_d^S(\mu^*(0), 0, 0)$ 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 naive 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 $a^R(0) < \mu^*(m) < \alpha^*(0)$. Therefore $U_d^S(\alpha^*(m), m, 0) > U_d^S(\alpha^*(0), 0, 0)$ and the type-0 dishonest Sender strictly benefits from sending m, a contradiction.

Hence, NITS holds in a message-monotone equilibrium in the perturbed game. This implies that only CS equilibria that satisfy NITS can be the limits of such equilibria as the non-strategic types vanish.

4.3.2 Costly Lying

Kartik (2005) studies the CS model augmented with lying costs for the Sender. A special case is as follows: assume that the message space and the type space are the same, so that $M = [0, 1]$. The game is identical to CS, except for Sender payoffs, which are given by $U^S(a, t) - kC(m, t)$, for $k > 0$. Assume that C is twice continuously differentiable, with $C_1(t, t) = 0$ and $C_{11}(m, t) > 0 > C_{12}(m, t)$, and normalize $C(t, t) = 0$. This is interpreted as an intrinsic cost of lying for the Sender, minimized by telling the truth, convex around the truth, and submodular: e.g., $C(m, t) = -(m - t)^2$. Plainly, when $k = 0$, the model is substantively equivalent to CS.

Kartik studies "monotone equilibria," which are pure strategy (perfect Bayesian) equilibria where the Sender's strategy, $\mu : [0, 1] \to [0, 1]$, is weakly increasing (message monotonicity, as in Chen) and the Receiver's strategy, $\alpha : [0, 1] \to \mathbb{R}$, is also weakly increasing (action or belief monotonicity). Note that when $k = 0$, there is no loss of generality with regards to equilibrium outcome mappings $T \to A$ in restricting attention to monotone equilibria.

Theorem 1 in Kartik (2005) says that as $k \to 0$, any convergent sequence of monotone equilibria must converge to a CS equilibrium that satisfies NITS. In fact, he shows that when k is small enough, all monotone equilibria must satisfy NITS, which implies that NITS must hold in any limit. The argument is as follows.

A basic implication of message monotonicity is that the set of types sending any given message is convex (possibly a singleton). 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 belief 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 belief 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, $\langle t_1^k, t_2^k \rangle$, that are pooling on some message $m_2 > 0$, with $U^S(\bar{a}(0, t_1^k), t_1^k) \approx U^S(\bar{a}(t_1^k, t_2^k), t_1^k)$.[8] 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 belief 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 cost, and elicit a weakly preferred action. Note also that because of the CS preference structure, there is a positive lower bound on how small $t_2^k - t_1^k$ can be.

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 belief monotonicity), and saving on lying cost: contradiction with equilibrium.

4.3.3 The Babbling Equilibrium

To further highlight the mechanism at work in Chen (2006) and Kartik (2005), let us focus on the one-step babbling equilibrium of CS. A general property of the equilibria in Chen's and Kartik's model, even if the respective perturbations are large, is that the highest message is always used. The reason is that there can be no unused messages "at the top," for such unused messages represent profitable deviations for all types close enough to 1: by belief monotonicity and lying costs in Kartik; by the behavior of the honest Sender and naive Receiver in Chen. Thus, if uninformative communication occurs, it must occur with all types sending the highest message. But if the uninformative CS outcome does not satisfy NITS, then types close to 0 have a profitable deviation to a sufficiently small message, since this induces an action close to $a^R(0)$ in Chen, and induces an action weakly lower than $\bar{a}(0, 1)$ while strictly saving on lying costs in Kartik.

A related mechanism is also present in Lo (2006), which we now briefly describe. Lo studies the CS model with restrictions on the strategy spaces, meant to describe limitations imposed by the use of a natural language.[9] She assumes that the message space is equal to the action space, which we assume are both $[a^R(0), a^R(1)]$. This does not restrict the set of equilibrium outcomes but makes it possible to associate messages with

[8] 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.

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

their "literal meaning." She models literal meaning by making two restrictions on the strategies available to the Receiver. The first restriction 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 $\alpha(a) = a$).[10] The second assumption requires the set of messages that induce a particular action be convex. The 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 (M), iterative deletion of weakly dominated strategies[11] 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 $m = a^R(1)$. This is similar to the equilibrium phenomenon noted earlier in Chen's and Kartik's models that types closes to 1 must send the highest message. 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.[12] It is then iteratively dominated for the Receiver to respond 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.

5 Applications

In this section, we apply NITS to two other models of one-dimensional cheap-talk with one-sided private information.[13]

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 show in this section that under certain conditions, a natural variation of NITS selects the informative equilibrium.

[10] This assumption is violated in the limiting equilibria of both the Chen and Kartik models.

[11] To be precise, all weakly dominated strategies for each player are deleted in each round.

[12] 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 $\alpha(m_1)$ to action $\alpha(m_2)$.

[13] In a model with two-sided private information, Chen (2007) shows that informative equilibria satisfy NITS and that the babbling equilibrium satisfies NITS only if it is the unique equilibrium 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 her 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 her 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 $s = 1$. 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).[14] 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 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 his 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. Let $a^P(t)$ be the action that the Proposer would take if the Chooser's type were known to be t.[15] 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 $t \in (0, 1)$ and $t > \underline{t}$ there always is such a type. 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.

[14] 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.

[15] The Proposer will offer his 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.

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 she would if she could reveal her type. Note that if the type-t Chooser reveals her type, then she 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 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:

$$t \text{ prefers } a^P(t) \text{ to } 0 \text{ implies } t' > t \text{ prefers } a^P(t') \text{ to } 0. \tag{12}$$

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 (12), it is not especially instructive. Instead, we simply assert that it holds when preferences are quadratic and the prior is uniform.[16]

Finally, we note that neologism-proof outcomes (Farrell, 1993) often fail to exist in Matthews's model.[17] 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.

[16] 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 (12) holds because $2t - a^P(t)$ is increasing.

[17] 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 he knows that the Chooser's type lies in T to the equilibrium payoffs.

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 $k \in (0,1]$.[18] Consequently, if a transfer is made with probability $1-a$, then $U^S(a,t) = (1-a)(1+ky) + a(t+k)$ 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

$$y^* := \frac{y}{k} + 1 - \frac{1}{k}. \tag{13}$$

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

$$\mathbb{E}[t|t < 1 - k(1-y)] \leq y^*. \tag{14}$$

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

[18] 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.

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.

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. (2006): "Perturbed Communication Games with Honest Senders and Naive Recievers," Mimeo, Arizona State University.

——— (2007): "Partially-informed Decision Makers in Games of Communication," Mimeo, 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.

Goltsman, M. and G. Pavolov (2006): "Mediated Cheap Talk," Mimeo, Boston University.

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.

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.