Information Transmission with Cheap and Almost-Cheap Talk*

Navin Kartik†
Department of Economics
University of California, San Diego

First version: November 2003
This version: January 2005

Abstract

Communication in practice typically occurs through multiple channels, not all of which permit costless misrepresentation of private information. Accordingly, I study a model of strategic information transmission based on Crawford and Sobel (1982), but allow for communication through both cheap talk and messages on a second dimension where misreporting is costly. Using a forward-induction refinement of sequential equilibrium, I characterize a class of equilibria with appealing properties. When misreporting costs are large, significant amounts of information can be transmitted in equilibrium. As the costs of misreporting become small, talk is almost-cheap, and the model is arbitrarily close to the pure cheap talk model. However, not all equilibria of the pure cheap talk model are limits of the equilibria with misreporting costs, and a simple condition is derived to determine which cheap talk equilibria are robust in this sense. I show that under a standard assumption, only the most-informative cheap talk equilibrium is robust. This provides a novel rationale for focussing on the most informative equilibrium of the cheap talk game, without invoking cooperative justifications such as the Pareto criterion.

Keywords: Cheap Talk, Costly Lying, Signaling, Refinements, Equilibrium Selection, Babbling, D1

J.E.L. Classification: C7, D8

*This paper is based on Chapter 1 of my dissertation at Stanford University. I am indebted to Doug Bernheim and Steve Tadelis for their generous support and advice. I thank David Ahn, Nageeb Ali, Vince Crawford, Peter Hammond, Cristóbal Huneeus, Mikko Packalen, Anjali Pai, Jeroen Swinkels, Bob Wilson, and especially Jon Levin, Ilya Segal, and Joel Sobel for helpful comments. The paper has benefitted from seminars at Brown, Columbia Business School, Harvard Business School, Kellogg, NYU, Stanford, Toronto, U-Penn, UC-Berkeley, UC-San Diego, UT-Austin, and Wisconsin. I gratefully acknowledge partial support from a John M. Olin Summer Research Fellowship and a Humane Studies Fellowship during this project.

†Email: nkartik@ucsd.edu; Web: http://econ.ucsd.edu/∼nkartik; Snail-mail: 9500 Gilman Drive, 0508, La Jolla, CA 92093-0508.

1 Introduction

The seminal work of Crawford and Sobel (1982, hereafter CS) on information transmission through cheap talk has been widely applied to problems in bargaining, organizational design, political science, monetary policy, and other areas.[1] The setting is a one-shot game where an informed party, the Sender, communicates his one-dimensional private information to an uninformed decision-maker, the Receiver, when their preferences over the decision may not entirely congruent.[2]

While the basic CS model has been extended in many directions,[3] an outstanding problem is that these theories are typically plagued by a multiplicity of equilibria. In particular, in CS, there generally is a range of equilibria that can be ranked in terms of their informativeness.[4] At one end of the spectrum is the most-informative equilibrium, and at the other end is the completely uninformative, babbling equilibrium. Unfortunately, there is no well-established criterion for selecting amongst these. Though the CS model is a signaling game, standard theory for equilibrium selection in signaling games (Cho and Kreps (1987) and Banks and Sobel (1987), based of the notion of stability in Kohlberg and Mertens (1986)) does not apply because the signals or messages are costless to the Sender.[5] Thus, there have been some alternative refinement approaches developed specifically for cheap talk games, mostly following ideas put forth by Farrell (1993). These have been only partially successful in general, and particularly unsuccessful when applied to CS. Notably, for the interesting cases, there is no CS equilibrium that survives criteria like *neologism-proofness* (Farrell 1993) or *announcement-proofness* (Matthews, Okuno-Fujiwara, and Postlewaite 1991); whereas all CS equilibria can survive *credible rationalizability* (Rabin 1990).[6]

The common approach taken in the applied literature is to focus on the most-informative equilibrium of the CS model. To the extent a theoretical justification is given,

[1]Respectively, see for example Farrell and Gibbons (1989a), Morgan and Stocken (2003), Dessein (2002), Matthews (1989), and Stein (1989).

[2]As a matter of convention, I treat the Sender as male and the Receiver as female throughout this paper.

[3]Gilligan and Krehbiel (1987, 1989), Austen-Smith (1993), and Krishna and Morgan (2001a, 2001b) study models with multiple senders; Battaglini (2002), Chakraborty and Harbaugh (2004), and Levy and Razin (2004) analyze multi-dimensional information; Farrell and Gibbons (1989b) develop an application with multiple receivers; Sobel (1985) and Morris (2001) analyze repeated cheap talk games; and Krishna and Morgan (2004) focus on multiple rounds of cheap in a one-shot game.

[4]Informally, the informativeness of an equilibrium refers to how much the Receiver learns about the Sender's private information after communication. A more-informative equilibrium ex-ante Pareto dominates a less-informative equilibrium provided that preferences satisfy a "similarity" assumption; see Condition M on p. 19. For the discussion in this Introduction, I assume the Condition holds.

[5]Any message from the Sender not sent in equilibrium can be interpreted by the Receiver in exactly the same way that some message that is sent in equilibrium is interpreted. Since messages are costless, no type of Sender would strictly prefer to send the out-of-equilibrium message over what he is supposed to in equilibrium. Standard refinements have no bite in this case. This logic implies in particular that every cheap talk equilibrium can be sustained with a strategy for the Sender that has full support over the set of available messages.

[6]A detailed discussion is postponed to Section 7.

it is that this is the ex-ante Pareto dominant equilibrium, and hence the one that players should be expected to coordinate on. There are at least two reasons why this is unsatisfactory. First, the Pareto criterion is a cooperative solution concept. Accordingly, its application in equilibrium selection for non-cooperative games is somewhat ad-hoc. Second, in some costly signaling games, Pareto dominant equilibria do not survive standard belief-based refinements (Banks 1991). This raises the possibility that a belief-based refinement for cheap talk games may also eliminate the ex-ante Pareto dominant equilibrium.

This paper approaches this issue from the position that communication in practice rarely consists of purely costless messages alone, especially in economically important situations.[7] As an illustration, suppose an environmental lobby group interacts with a politician. The lobby is the Sender, privately informed about the effect of pollution on the environment; the politician is the Receiver, who must choose the level to cap emissions for some industry. There are various ways in which the lobby can potentially transmit information. It might be required to submit a report to the politician, detailing the relevant evidence, testimonies from experts, etc. This would not be pure cheap talk: it is natural to think that on this dimension, misreporting its information is costly. The costs may stem from evidence fabrication (an expert has to be bribed in order to lie), exogenous auditing probability (the politician may have an independent body check the veracity of the report), reputational effects (public opinion if the deception is discovered), or even internal "ethics" or "psychological costs". On the other hand, pure cheap talk is also typically available: in addition to the the report, the lobby could make various verbal remarks to the politician, or potentially reveal something by submitting the report sooner rather than later, etc.[8] These are cheap talk in the usual sense that there is no exogenous cost of sending one signal rather than another on these dimensions.

As another example, consider a worker who possess some information about future demand forecast and must communicate with his manager. On the one hand, if the worker literally misreports about the forecast, he faces some risk of being contradicted by another source, or the manager randomly verifying the information, and so forth. This generates a cost of explicit lying. On the other hand, he could also make pronouncements that are not literally about the value of the forecast, such as "I've heard that the market might be slower than expected," or "Remember how two years ago we did so much better than anticipated?", etc. These are not extrinsically costly (non-verifiable, no evidence fabrication needed), but could nonetheless convey information.

More generally, it is often the case that there are dimensions of communication that have intrinsic lying or *misreporting costs*, while there are other dimensions that are costless. Therefore, I present a model where a Sender is privately informed about a one-dimensional

[7] This point was also noted by Austen-Smith and Banks (2000) in a vivid and striking example. Their focus is different from this paper, however.

[8] In principle, the domain of such costless messages is vast. For an extreme example, observe that whether the lobbyist says "hello" rather than "hi" when delivering the report could (in equilibrium) reveal different things.

variable but communicates to an uninformed Receiver via signals of *two* kinds: the first, which I refer to as a *report*, entails a cost of misreporting his private information that increases in the degree of misreporting; the other, which I call a *message*, is costless and pure cheap talk. To facilitate a comparison with CS, I assume that the basic payoffs to both players are simply the CS payoffs, and then augment an added cost to the Sender stemming from any misreporting with the costly signal. Parameterizing the relative magnitude of these costs by a positive scalar k, it is intuitive that we are back in the case of CS pure cheap talk when $k = 0$. The case of k close to 0 is referred to as *almost-cheap talk*.

The model is subject to multiple equilibria just as CS and most signaling models are. However, due to the presence of the costly report, I am able to use a variant of the D1 refinement for signaling games to characterize a class of equilibria that are appealing to focus on.[9] Existence of equilibria in this class is proven for any k. When k is large, there are equilibria with full separation through the costly report below some interior cutoff type, and partial separation through cheap talk above this cutoff type. For small enough k (or almost-cheap talk), generically all equilibria in the class I study feature complete pooling on the costly report, and partial separation through cheap talk. As separation through cheap talk alone is analogous to that of CS, it follows that at least one CS equilibrium outcome is generically an equilibrium outcome of my model for small k. I derive a simple condition that determines whether a particular CS equilibrium outcome is an equilibrium of the current model with almost-cheap talk: it only needs to be checked whether the lowest type of Sender would prefer to be pooled as in the CS equilibrium rather than separate himself if he could. If pooling is preferred, the CS equilibrium remains an equilibrium once talk is almost-cheap; if separation is preferred, the CS equilibrium fails to be an equilibrium for any magnitude of misreporting cost, and moreover, is not the limit of any sequence of equilibria (in the class I study).

The intuition behind this relies on the nature of beliefs that are required to hold in the forward-induction equilibria. Regardless of how small misreporting costs are, so long as they are positive, the belief refinement implies that the Receiver must put probability one on the lowest type of Sender upon receiving an out-of-equilibrium costly report which is lower than all in-equilibrium costly reports. Generically, once misreporting costs are sufficiently small, all equilibria feature complete pooling on the costly reporting dimension with all types sending the highest available costly report. Hence, there are unused costly reports in equilibrium that allow the lowest type to separate himself, should he so desire. This implies that the lowest type must prefer its equilibrium payoff to what he would get by separating itself.

Under a standard regularity condition,[10] I show that only the most-informative CS equilibrium is robust in the sense that it is the limit of forward-induct equilibrium outcomes

[9] The particular refinement I use is the *mD1* criterion, introduced by Bernheim and Severinov (2003), extending the work of Cho and Kreps (1987) and Banks and Sobel (1987).

[10] The "monotonicity" condition of CS, Condition M defined on p. 19 of this paper.

of my model with misreporting costs. This paper therefore provides a theory of equilibrium selection for the CS framework. It justifies the applied practice of focussing on the most-informative equilibrium in CS.

The rest of the paper is structured as follows. I discuss the related literature in the following section. Section 3 lays out the model, and Sections 4 and 5 characterize equilibria. I analyze the behavior of equilibria as k gets small in Section 6, and present the CS equilibrium selection results. I discuss the theory more broadly in Section 7, including its relationship to other cheap talk refinements, and some extensions. A brief conclusion follows in Section 8.

2 Related Literature

This work is connected to a few different strands of literature. The most closely related paper is that of Bernheim and Severinov (2003). Although their setting and application is very different, the formal structure of their model is related to this paper in some respects. The equilibrium refinement I use was introduced by them. There are important differences however, foremost that much of the analysis here concerns behavior as the cost of the discriminatory signaling goes to 0, which is not the focus of their work. Moreover, their model is one of multidirectional signaling that results in a "central pool", whereas my model is one of unidirectional signaling that results in pool at the top end of the type space.

In the pure cheap talk literature, the most relevant work is that of Crawford and Sobel (1982). There is a large literature on refining cheap talk equilibria. While the most famous are arguably those of Farrell (1993), Matthews, Okuno-Fujiwara, and Postlewaite (1991), and Rabin (1990), the closest ones in spirit to the approach taken here are the "perturbation methods" studied by Blume (1994) and Blume (1996).

With respect to the pure discriminatory signalling literature, the refinement approaches of Cho and Kreps (1987) and Banks and Sobel (1987) are the bases for the particular refinement I adopt. Cho and Sobel (1990) were the first to obtain incomplete separation in the manner obtained here: pooling at the top of the type space and separation at the bottom with respect to the costly signal.

Finally, Austen-Smith and Banks (2000) also present a model with cheap talk and signaling, but they study money-burning as opposed to discriminatory signaling. Moreover, their focus is on how the set of outcomes is expanded vis-a-vis pure cheap talk, rather than refining cheap talk equilibria.

3 Model

There are two players, a Sender (S) and a Receiver (R). The Sender has private information summarized by his type $t \in T \equiv [0,1]$, which is drawn from a differentiable probability distribution $F(t)$, with density $f(t) > 0$ for all $t \in T$. After privately observing his type

t, S sends R a signal pair (r, m) consisting of a report, $r \in T$, and a message, $m \in M$, where M is an arbitrary uncountable space. R then takes an action, $a \in \mathbb{R}$. The report r is payoff relevant for S (but not for R) with misreporting cost given by $C(r, t)$, while the message m is pure cheap talk and thus payoff irrelevant to both players. The payoff for R is given by $V(a, t)$, and the payoff for S is given by $U(a, t) - kC(r, t)$, with $k > 0$ a measure of the magnitude of misreporting costs. All aspects of the game except the value of t are common knowledge.

Throughout, the following assumptions on payoffs are maintained. The functions $U(a, t)$ and $V(r, t)$ are twice continuously differentiable. Using one or two subscripts to denote first and second derivatives respectively, $U_{11} < 0 < U_{12}$ and $V_{11} < 0 < V_{12}$, so that both the Sender and the Receiver prefer higher actions given higher types. For any t, there exists $a^R(t)$ and $a^S(t)$ respectively such that $V_1\left(a^R(t), t\right) = U_1\left(a^S(t), t\right) = 0$, with $a^S(t) > a^R(t)$. That is, the most-preferred actions are well-defined for both players, and the Sender prefers higher actions than the Receiver.[11] The assumptions on U and V imply that for $i \in \{R, S\}$, $a_1^i(t) > 0$. Finally, $C(r, t)$ is twice continuously differentiable, with $C_{11} > 0 > C_{12}$, so that the marginal cost of misreporting is increasing as the report gets further away from the true type, and higher types prefer higher reports. Consistent with C being misreporting costs, $C_1(t, t) = 0$ for all $t \in T$, so that the cheapest report for any type is the truth.

A few aspects of the model are worth emphasizing. First, note that the report r is a *discriminatory* signal, in the sense that the cost varies with type, and not non-discriminatory or *money-burning*. Second, if $k = 0$, the model essentially collapses to that of CS, except for the added report, r. So long as the cheap talk message space M is rich enough (e.g. uncountable, as was assumed), the extra dimension of signaling adds nothing when $k = 0$. However, when $k > 0$, as is assumed, the costly report r can potentially play an important role in information transmission. Third, the Sender's report is assumed to lie in the type space, T. While I discuss in Section 7.2 what happens if this is relaxed, the conceptually crucial point is that there is an ordering on the report space such that one can meaningful talk about the cost of misreporting, and how this varies across types. The most natural modelling choice is simply that of $r \in T$. Returning to the motivating example, the sales agent's type would be his estimate of demand. It is sensible to require that in his report to the manager, he must specify some number in the commonly known range of demand, which is normalized to $[0, 1]$.[12] Of course, one could also assume that the cheap talk message space is T, and doing so would not change the analysis. It is useful however to allow M to be any arbitrary rich space, to emphasize that these messages are just informal communication, and need not have an ordering or any particular relationship with the type space. Recall that these messages may be gestures, tone of voice, etc. Lastly,

[11] What is important is that there is no t such that $a^R(t) = a^S(t)$. Given this, $a^R(t) > a^S(t)$ is without loss of generality.

[12] I discuss what happens when the state space is unbounded in an Appendix.

I have assumed that each type's ideal report is the truth, which seems like the most natural assumption given the interpretation of costs as stemming from as misreporting. However, the analysis remains essentially the same under much weaker conditions: it is sufficient if cost minimizing reports are well-defined for each type, and increasing with type. This is discussed further in Section 7.2.

A clarification on terminology in what follows: I use the term *report* or *costly signal* to refer to r, the term *message* to refer to the cheap talk message m, and the unqualified term *signal* to refer to a report-message pair (r, m). The Receiver's decision a is termed *action.*

The basic equilibrium concept is Sequential Equilibrium (Kreps and Wilson 1982), formulated the standard way for infinite signaling games (see for example Manelli (1996)). The Sender's strategy is given by a pair of functions (ρ, μ) where $\rho : T \to T$ defines the Sender's report or costly signal as given his type, and $\mu : T \to M$ defines the Sender's cheap talk message given his type. Denote the posterior beliefs of the Receiver given r and m by the cumulative distribution $G(t \mid r, m)$. The Receiver's strategy is denoted $\alpha : T \times M \to \mathbb{R}$. By the strict concavity of V, it is clear that the Receiver will never optimally play a mixed strategy. The restriction to a pure message strategy for the Sender is also without loss of generality since messages are costless. That the Sender uses a pure strategy on the costly signal (the report) is a restriction that is common in the literature. Equilibrium requires the Sender to be playing optimally given the Receiver's strategy, and that the Receiver play optimally with respect to his beliefs, which must be formed according to Bayes' Law for every signal that is sent in equilibrium.

It is worthwhile to emphasize a technical issue at this point.[13] The model is a signaling game with a continuous signal and type space. As such, without the cheap talk dimension, there are no general existence results on sequential equilibria. Manelli (1996) proves that for a *cheap talk extension* of a continuous signaling game, sequential equilibria exist.[14] The current model can be thought of as the cheap talk extension of a game where the Sender only has available the costly report (r). In that game, it is not known whether sequential equilibria exist in general; in the current game, on the other hand, Manelli's (1996) results assure it.[15] In what follows, I prove existence in and characterize a subset of all sequential equilibria that satisfy a forward-induction refinement along the lines of Cho and Kreps's (1987) *D1* criterion.[16] The construction I use relies on the cheap talk

[13] I thank Jeroen Swinkels for a very helpful discussion on this matter.

[14] Informally, given a signaling game, its cheap talk extension is defined as an augmented signaling game where the Sender's signal space is the Cartesian product space of the original signal space and the Receiver's action space, but everything else including payoffs stay the same. That is, in the cheap talk extension, the Sender sends not only a signal of his type, but also a payoff-irrelevant "recommendation" to the Receiver.

[15] See also Jackson, Simon, Swinkels, and Zame (2002) for an analysis of the role of cheap talk in assuring equilibrium in a broad class of incomplete information games with infinite type and action spaces.

[16] See also Banks and Sobel (1987). Manelli (1997) and Ramey (1996) prove existence of forward-induction equilibria for certain infinite signaling games that do not include the current model.

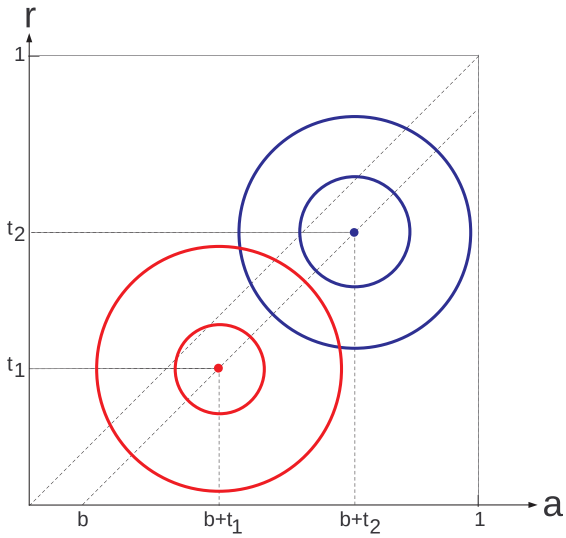

Figure 1: Elliptical Indifference Curves

dimension, and indeed, without cheap talk, it is not hard to construct generic examples where equilibria fail to exist in the class I consider. While I believe that cheap talk is economically viable in many cases, a different interpretation of the model would be to view the cheap talk dimension as purely a theoretical construct to study equilibria. Under this interpretation, the pure cheap talk game of CS can be viewed as the limit of costly misreporting games where the Sender has available *only* the costly report, but the formal analysis is conducted using cheap-talk extensions of these games.[17]

4 No Separating Equilibria

In this section, I show that there cannot be fully separating equilibria.[18] Before proceeding to the analysis, it is worth pointing out the model does *not* have a single crossing property (SCP) in (a, r)-space, unlike most traditional signaling models. This is because indifference curves in (a, r)-space are elliptical, and hence either cross twice or not at all. This is illustrated in Figure 1, where each type's ideal action is $a^S(t) = t + b$ (for some exogenously given bias parameter $b > 0$). Hence the most preferred point for type t in (a, r)-space is $(t + b, t)$.

[17]Trivially, the cheap talk extension of the CS model has the same set of equilibrium outcomes, i.e. mappings from Sender types to Receiver actions, as CS.

[18]In many standard signaling games, fully separating equilibria are not only guaranteed to exist, but are often isolated using refinement criteria (Mailath 1987, Ramey 1996). However, these arguments typically require a sufficiently large signal space and the single crossing property.

Note that in a fully separating equilibrium, even though multiple types may be using the same report, they are correctly identified by R, and hence any type t induces an action of $a^R(t)$. We start with an important Lemma which says that if there is full separation below some type $\bar{t}$, then no non-zero type $t \leq \bar{t}$ can be playing his ideal report. The intuition is simple: if there were such a type, call it $\hat{t}$, a slightly smaller type could imitate $\hat{t}$, and would suffer at most only a second-order loss through misreporting, but benefit by a first-order gain in the induced action a.

Lemma 1. *In any equilibrium, if the Sender plays* (ρ, μ) *and is separating up to and including some type* $\tilde{t}$*, then for all* $t \in (0, \tilde{t}]$*,* $\rho(t) \neq t$*.*

The proof (and all subsequent proofs not in the text) is relegated to the Appendix. With this Lemma in hand, I prove that there is no separating equilibrium.

Proposition 1. *There is no separating equilibrium.*

The gist of the argument is as follows. By the Lemma, no non-zero type can be using his ideal report, and in particular, the highest type cannot be using the highest report. So if there is some type t using a report $\rho(t) > t$, then the report strategy must discontinuously jump from above the 45° line (which graphs the ideal report for each type) to below it, at some $t' > t$. The main part of the proof shows that this cannot be the case. Suppose $\rho(t') < t'$. Then for some type t'' slightly below t', the gain in utility for either type by being perceived as type t' rather than type t'' is of the order of $(t'' - t')^2$. On the other hand, the change in misreporting costs is of the order of $(\rho(t'') - \rho(t'))(t'' - t)$, which is of the order of $(t'' - t)$ since $\rho(t'') - \rho(t')$ is bounded away from 0 by hypothesis. Since t' values higher reports more than t'', if t' is willing to play $\rho(t') < \rho(t'')$, then so will type t'', contradicting separation. Thus, the report strategy lies strictly below the 45° line for all non-zero types. It is easily verified that it must be continuous everywhere and hence weakly increasing around type 0. But then some small type t would prefer to imitate a slightly higher type t', since he gains in the induced action without losing on the misreporting cost.

The Proposition implies that there must be some pooling in equilibrium.[19] Hence the presence of costly signalling does not completely alleviate the information asymmetry.

5 Pooling Equilibria

As is typically the case with signaling games, the current model may have multiple equilibria. Accordingly, I use a refinement that is in the spirit of the well-known D1 criterion of Cho and Kreps (1987).

[19] The proof of the Proposition makes use of the fact that the ideal report for the highest type is in fact the highest report. This might suggest that requiring the report space to be just the type space is a severe restriction. However, more generally, for any fixed upper bound on the report space, once k is small enough, there cannot be a fully separating equilibrium. I discuss this further in Section 7.2.

5.1 The Monotonic D1 Refinement

The specific criterion I use is the *monotonic D1* (mD1) criterion, introduced by Bernheim and Severinov (2003) in a different context, but in a formally related model. The basic idea is the same as the D1 criterion, which says that upon observing an out-of-equilibrium signal, the Receiver should not believe it is a type t if there is some other type t' who would strictly prefer to deviate for any response from the Receiver that type t would weakly prefer to deviate for. In addition, the monotonicity requirement is that higher types use higher signals. Let us develop these concepts formally.

Definition 1. An equilibrium is report-monotone if (i) $\rho(t)$ is weakly increasing, and (ii) $G(t \mid r, m)$ is weakly decreasing in r for all t, m.

The first part of the definition is straightforward. It says that higher types must send weakly higher reports. In models where the SCP is satisfied, this is a property that *must* hold in any sequential equilibrium. This guarantees that R's beliefs when seeing a report r must first order stochastically dominate (FOSD) her beliefs upon seeing report $r' < r$, whenever r and r' are both *on the equilibrium* path. Part (ii) of the definition requires this to hold for *off the equilibrium path* reports as well. It is worth mentioning that in standard models, any sequential equilibrium is outcome-equivalent to one where beliefs satisfy this property on and off the equilibrium path (Cho and Sobel 1990). A useful and straightforward consequence of report monotonicity is that R's strategy must be increasing in the report she hears.

Lemma 2. *In a report-monotone equilibrium, $\alpha(r, m)$ is weakly increasing in r for all m.*

Proof. Fix a message m. By a well-known property of FOSD, $\int V_1(a,t)\, d\Phi(t,r,m)$ is increasing in r since V_1 is increasing in t ($V_{12} > 0$). Therefore, $\int V(a,t)\, d\Phi(t,r,m)$ has increasing differences in a, t. By Topkis' Theorem, the maximizers are weakly increasing. □

One other piece of notation is required before the mD1 criterion is defined. We will need to refer to the highest or lowest action played in response to reports lower or higher than a given report. Formally, let

$$\xi_l(r) \equiv \begin{cases} \sup_{t:\rho(t)<r} \alpha(\rho(t), \mu(t)) & \text{if } \exists t \text{ s.t. } \rho(t) < r \\ a^R(0) & \text{otherwise} \end{cases}$$

$$\xi_h(r) \equiv \begin{cases} \inf_{t:\rho(t)>r} \alpha(\rho(t)\, \mu(t)) & \text{if } \exists t \text{ s.t. } \rho(t) > r \\ a^R(1) & \text{otherwise} \end{cases}$$

For an out-of-equilibrium report r' such that some report $r < r'$ ($r > r'$) is sent in equilibrium, $\xi_l(r')$ ($\xi_h(r')$) gives the highest (lowest) action taken by the Receiver in response to an equilibrium report lower (higher) than r. If r' is such that there is no report $r < r'$ $(r > r')$ sent in equilibrium, then $\xi_l(r')$ $(\xi_h(r'))$ just specifies the lowest (highest) rationalizable action for the Receiver. Obviously, for all r (on or off the equilibrium path), $a^R(0) \leq \xi_l(r) \leq \xi_h(r) \leq a^R(1)$; this is just a consequence of the fact that sequential equilibrium requires R's action to be optimal for *some* belief.

The following is the restriction on beliefs introduced by Bernheim and Severinov (2003), restated for the current model.

Definition 2. An equilibrium satisfies the monotonic D1 criterion if

1. It is report-monotone.

2. For any off-the-equilibrium report r, if there is a nonempty set $\Omega \subseteq T$ such that for each $t \notin \Omega$, there exists some $t' \in \Omega$ such that for all $a \in [\xi_l(r), \xi_h(r)]$,

$$\begin{array}{rcl} U(a,t) - kC(r,t) & \geq & U(\alpha(\rho(t), \mu(t)), t) - kC(\rho(t), t) \\ & \Downarrow & \\ U(a,t') - kC(r,t') & > & U(\alpha(\rho(t'), \mu(t')), t') - kC(\rho(t'), t') \end{array}$$

Then for all m, $Supp\, G(t \mid r, m) \subseteq \Omega$.

The first part of the definition is straightforward. If we replace $a \in [\xi_l(r), \xi_h(r)]$ with just $a \in [a^R(0), a^R(1)]$ then part 2 of the definition would be precisely the D1 criterion of Cho and Kreps (1987). However, if the Receiver's beliefs respect monotonicity, then for any out of equilibrium report r, and any message m and type t, $G(t \mid r, m) \geq \sup_{t:\rho(t)<r} G(t \mid \rho(t), m)$, and optimality requires $\alpha(r, m) \geq \xi_l(r)$. Similarly one sees that $\alpha(r, m) \leq \xi_h(r)$. Accordingly, part 2 of the definition above applies the idea behind the D1 criterion on the restricted action space $[\xi_l(r), \xi_h(r)]$. That is, it requires that for some out-of-equilibrium report r, if there some type t' who would *strictly* prefer to deviate for any action $a \in [\xi_l(r), \xi_h(r)]$ that a type t would *weakly* prefer to deviate for, then R exclude type t from the support of her beliefs.

For the main part of the paper, I restrict attention to mD1 equilibria. They are appealing for the same reasons that D1 equilibria are appealing in standard signaling games. The added monotonicity requirement imposed by mD1 is reasonable and desirable in the current context because higher types prefer higher reports.

Remark 1. The mD1 criterion is at least weakly stronger than the D1 criterion, so that the set of mD1 equilibria is a weak subset of the set of D1 equilibria. However, in standard "monotonic" signaling games (Cho and Sobel 1990) where all types of the Sender wish to be perceived as the highest type, and a single-crossing property in Sender's signal-Receiver's action space holds, the set of mD1 equilibria is identical to the set of D1 equilibria.[20] The

[20]This follows from Cho and Sobel's (1990, Lemma 4.1) characterization of D1 equilibria in this class of games.

problem is that for the current model, which is not monotonic in the above sense, the D1 criterion does not restrict the set of sequential equilibrium outcomes in general.

Remark 2. As with all standard signaling game refinements, the mD1 criterion does not help restrict the set of sequential equilibrium outcomes when $k = 0$. To see this, consider a sequential equilibrium when $k = 0$. Given that M is uncountable and $k = 0$, there is an essentially equivalent sequential equilibrium (i.e. one that induces the same mapping from types to actions) where all types send the same report, call it r^*, and use possibly different cheap talk messages. I claim that this equilibrium can be supported by strategies that satisfy mD1. As in CS (see Lemma 3 on p. 14 of this paper), there can only be a finite number of actions induced in equilibrium when $k = 0$; hence the equilibrium can be supported with finitely many distinct cheap talk messages. Denote the highest and lowest actions induced in equilibrium by a_h and a_l respectively with corresponding messages m_h and m_l. For any m and t, define for all $r < r^*$, $G(t \mid r, m) \equiv G(t \mid r^*, m_l)$, and for all $r > r^*$, $G(t \mid r, m) \equiv G(t \mid r^*, m_h)$. Then for all m, if $r < r^*$, $\alpha(r, m) = a_l$ and if $r > r^*$, $\alpha(r, m) = a_h$. Clearly these strategies continue to form a sequential equilibrium that supports the same outcome as the original equilibrium. Since $\rho(t) = r^*$ for all t, the given specification of G makes the equilibrium report monotone; hence part 1 of the mD1 criterion is satisfied. To see that part 2 also is, suppose by way of contradiction that it is not for some out-of-equilibrium report $r > r^*$ (the argument is analogous for $r < r^*$). First observe $\xi_l(r) = a_h$ and $\xi_h(r) = a^R(1)$. Given that part 2 of mD1 is being violated, there must be either (i) a type t in the support of $G(\cdot \mid r, m)$ for some m who strictly prefers the equilibrium action he induces to a_h; or (ii) a type t' who strictly prefers a_h to the equilibrium action he induces. But (i) but cannot be true because then t does not induce a_h in equilibrium, and since $G(\cdot \mid r^*, m_h)$ is derived by Bayes' Rule ((r^*, m_h) is an on-the-equilibrium-path signal), t cannot be in the support of $G(\cdot \mid r^*, m_h)$, and by definition, $G(t \mid r, m) = G(t \mid r^*, m_h)$. Yet (ii) cannot be true either because then t' would have a strictly profitable deviation to (r^*, m_h) over his equilibrium signal, which contradicts equilibrium.

5.2 Characterization

In this section, mD1 equilibria are characterized; existence is then established in Section 5.3. I first state the main theorem and discuss it. Thereafter, the remainder of the section is devoted to presenting the key steps in proving the Theorem.

Let ρ^* be the unique solution to the following Initial Value ODE Problem. (That there is a unique solution is established in Lemma 6.)

$$\rho'(t) = \frac{U_1\left(a^R(t), t\right) a_1^R(t)}{kC_1(\rho(t), t)}, \qquad \rho(0) = 0 \tag{DE}$$

As discussed later, ρ^* is strictly increasing, and accordingly there is a unique solution to

$\rho^*(t) = 1$, which I denote by $\bar{t}$. It is established later that $\bar{t} < 1$. Lastly, define for any $t'' \leq t'$, $\overline{a}(t'', t')$ to be the optimal action for the Receiver if the only information she has is that the Sender's type lies in $[t'', t']$. That is,

$$\overline{a}(t'', t') \equiv \begin{cases} \arg\max \int_{t''}^{t'} V(a, t)\, dF(t) & \text{if } t' > t'' \\ a^R(t') & \text{if } t' = t'' \end{cases}$$

Theorem 1. *In any mD1 equilibrium,* $(\hat{\rho}, \hat{\mu}, \hat{\alpha})$*, there exists some* $\hat{t} \in [0, \bar{t}]$ *and a partition of* $[\hat{t}, 1]$ *given by* $\{t_0 = \hat{t}, t_1, ..., t_{J-1}, t_J = 1\}$ $(J \geq 1)$ *such that*

i. $\forall j = 1, ..., J-1,$
$$U(\overline{a}(t_j, t_{j+1}), t_j) = U(\overline{a}(t_{j-1}, t_j), t_j) \tag{A}$$

ii. If $\hat{t} > 0$ *then*
$$U(a^R(\hat{t}), \hat{t}) - kC(\rho^*(\hat{t}), \hat{t}) = U(\overline{a}(\hat{t}, t_1), \hat{t}) - kC(1, \hat{t}) \tag{CIN}$$

iii. If $\hat{t} = 0$ *then*
$$U(a^R(0), 0) - kC(0, 0) \leq U(\overline{a}(0, t_1), 0) - kC(1, 0) \tag{ZWP}$$

and there is a set of J *distinct messages* $\{m_0, \ldots, m_{J-1}\}$ *such that strategies satisfy*

a. $\forall t < \hat{t}$, $\hat{\rho}(t) = \rho^*(t)$; $\forall t \in (\hat{t}, 1]$, $\hat{\rho}(t) = 1$; $\hat{\rho}(\hat{t}) \in \{\rho^*(\hat{t}), 1\}$; *if* $\hat{t} = 0$ *and* (ZWP) *holds with strict inequality then* $\hat{\rho}(0) = 1$

b. $\forall j = 0, ..., J-1,$

(b.1) $\forall t \in (t_j, t_{j+1})$, $\hat{\mu}(t) = m_j$ $\quad (m_j \neq m_n\ \forall n \neq j)$

(b.2) $\hat{\alpha}(1, m_j) = \overline{a}(t_j, t_{j+1})$

c. $\forall t < \hat{t}$, $\hat{\alpha}(\hat{\rho}(t), \hat{\mu}(t)) = a^R(t)$

Conversely, for any $\hat{t} \in [0, \bar{t}]$ *and* $\{t_0 = \hat{t}, t_1, ..., t_{J-1}, t_J = 1\}$ *that satisfy (i)-(iii) above, there is an mD1 equilibrium with full separation below* $\hat{t}$*, and bunching above* $\hat{t}$ *according to the given partition, with strategies satisfying (a)-(c).*

In words, the Theorem says that any mD1 equilibrium can be described by a *cutoff type*, $\hat{t}$, and a sequence of *boundary types* $\{t_0 = \hat{t}, t_1, ..., t_J = 1\}$ $(J \geq 1)$ such that three conditions are satisfied: (i) for any $j \in \{1, ..., J-1\}$, the boundary type t_j is indifferent between being perceived as a member of $[t_{j-1}, t_j]$ or a member of $[t_j, t_{j+1}]$ (this is condition A, for *arbitrage*, following CS); (ii) if the cutoff type $\hat{t}$ is strictly interior, then $\hat{t}$ is indifferent between being perceived as a member of $[\hat{t}, t_1]$ and incurring the cost of report 1, or separating himself and incurring the cost of report $\rho^*(\hat{t})$ (this is condition CIN, for

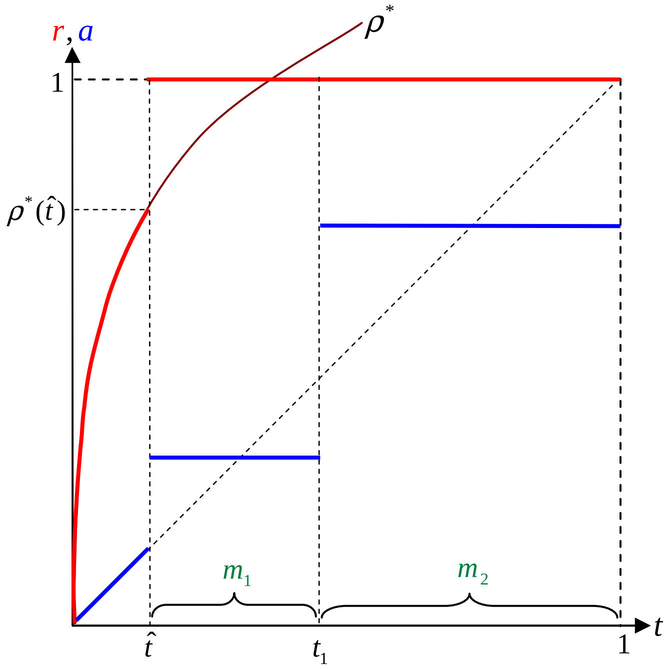

Figure 2: An mD1 equilibrium

cutoff indifference); (iii) if the cutoff type is 0 then type 0 weakly prefers being perceived as a member of $[0, t_1]$ and incurring the cost of report 1 to separating himself and incurring the cost of report 0 (this is condition ZWP, for *zero weak preference*). Conversely, any $\hat{t}$ and $\{t_0 = \hat{t}, t_1, ..., t_{J-1}, t_J = 1\}$ that satisfy these three conditions characterizes an mD1 equilibrium.

In any mD1 equilibrium, all types below $\hat{t}$ are separating themselves using report strategy $\rho^*(t)$, and all types above $\hat{t}$ are pooling on report 1, but separating themselves according to the given partition of $[\hat{t}, 1]$ by using cheap talk messages. Figure 2 illustrates the structure of an mD1 equilibrium with a strictly positive cutoff type and two distinct cheap talk messages sent in equilibrium. The function ρ^* is plotted as the brown line, the equilibrium reporting strategy is the red line, and the actions induced by each type is the blue line. The figure illustrates that types $t < \hat{t}$ separate themselves by reporting $\rho^*(t)$, whereas all type $t > \hat{t}$ report 1. Nonetheless, there is segmentation within the upper pool through cheap talk: types $t \in (\hat{t}, t_1)$ send cheap talk message m_1, whereas types $t \in (t_1, 1)$ send m_2 ($m_2 \neq m_1$). Assuming that the Receiver's optimal action is the expectation of the Sender's type and that the Sender types are distributed *ex-ante* uniformly over $[0, 1]$, the induced mapping from types to actions is the 45° line for $t < \hat{t}$, $\frac{\hat{t}+t_1}{2}$ for types $t \in [\hat{t}, t_1)$, and $\frac{1+t_1}{2}$ for types $t \in [t_1, 1]$.

5.2.1 Proving the Theorem

We begin the analysis with some preliminary observations. Report monotonicity has a direct implication on what set of types could be pooling on any report.

Observation 1. *In a report-monotone equilibrium, the set of types using the same report must be connected.*

That is, given a report r, if one defines $t_l(r) \equiv \inf\{t : \rho(t) = r\}$ and $t_h(r) \equiv \sup\{t : \rho(t) = r\}$, then every type $t \in (t_l(r), t_h(r))$ sends report r. It is important however to recognize that the set of types using r need not be indistinguishable to the Receiver, since they could potentially send different cheap talk messages. Accordingly, a *pool* refers to the set of types using the same report, whereas a *bunch* refers a set of types using both the same report and the same message. Obviously, the set of types bunching on some (r, m) must be a subset of the set of types pooling on r, hence contained within $[t_l(r), t_h(r)]$. By the supermodularity of U and that messages are costless, two facts follow: first, within any pool, there can be only a finite number of bunches; and second, bunches must be connected. The reasoning is identical to that of Crawford and Sobel (1982, Theorem 1), since within a pool, we are basically back in the CS world where any separation can only be achieved via cheap talk. To state this formally, let us say that an action a is *elicited* by report r if there exists some message m such that $\alpha(r, m) = a$. Furthermore, say that two equilibria are *essentially equivalent* if they induce the same mapping from types to reports and actions.

Lemma 3. *In any report-monotone equilibrium, every report elicits only a finite number of actions. Moreover, the equilibrium is essentially equivalent to a report-monotone equilibrium where bunches are connected.*

Accordingly, I focus without loss of generality on equilibria where bunches are connected. It is worth drawing out one implication of the Lemma that is often used in the sequel: if there is a non-trivial pool on a report in a mD1 equilibrium, no type in the pool can be separating, and in particular, there is no separation at the top or the bottom of the pool. This is an immediate consequence of the finiteness of elicited actions, since by continuity of U, the set of types who most prefer a given action over all others amongst a finite set of actions cannot be a singleton. Accordingly, a pool on a report r consists of $J \geq 1$ bunches that can be described by a partition $\{t_0 = t_l(r), t_1, ..., t_{J-1}, t_J = t_h(r)\}$ such that all the types in a given bunch, $t \in (t_j, t_{j+1})$, send the same cheap talk message, and for any $n \neq j$, types $t \in (t_n, t_{n+1})$ send a different cheap talk message.

The key result towards characterizing mD1 equilibria is the following Lemma, which says three things: first, that if there is pooling on some $r_p < 1$, then there are some reports immediately above r_p that are unused in equilibrium; second, upon seeing such an unused report, the Receiver must believe that it was sent by type $t_h(r_p)$; and third, seeing any unused reports that are lower than every used report must induce the Receiver to believe it was sent by the type 0.

Lemma 4. *In any mD1 equilibrium,*

1. *If there is pooling on some report $r_p < 1$, then there exists some $\theta(r_p) > 0$ such that reports $r \in (r_p, r_p + \theta(r_p))$ are unused in equilibrium.*[21]

2. *For all such $r \in (r_p, r_p + \theta(r_p))$, and any message m, $\alpha(r, m) = a^R(t_h(r_p))$.*

3. *If $\rho(0) > 0$, then for all m and $r < \rho(0)$, $\alpha(r, m) = a^R(0)$.*

Let us discuss the intuition behind the result. Consider the first part. If $\rho(1) = r_p$, then it is obvious, so the interesting case is when $\rho(1) > r_p$. I claim that to deter the highest type in the pool on r_p — which I'll call simply t_h to save notation — from mimicking a slightly higher type, there must be a discontinuous jump in the report strategy at t_h. If this were not the case, then the types immediately above t_h must be separating themselves, since by definition, t_h is the highest possible type using r_p. By the fact that no type in a pool can be separating, type t_h is inducing an action strictly smaller than $a^R(t_h)$. But then, by mimicking some type $t_h + \varepsilon$, type t_h would get a gain in utility from the induced action that is bounded away from 0, and only suffer an arbitrarily small loss in misreporting cost, by the continuity of the reporting strategy. This contradicts t_h optimally pooling on r_p.

Now turn to the second part of the Lemma. I argue that the mD1 criterion requires the Receiver to put probability 1 on type t_h for all such reports r. To see how this works, let a_- be the highest action elicited by any type in the pool on r_p, and a^+ be the lowest action elicited in equilibrium by types above t_h. If type t_h would want to strictly deviate from his equilibrium play for every response $a \in [a_-, a^+]$ to report r that any other type t would want to weakly deviate for, then mD1 requires R to place probability 1 on t_h upon seeing r. Consider some type $t < t_h$. In equilibrium, t induces an action weakly lower than that induced by t_h, and weakly lower than a_-. Moreover, t's preferred action and t's preferred report are both strictly lower than t_h's. The supermodularity of U and the submodularity of C imply that if t would prefer to deviate, then so would t_h. An analogous argument works for $t > t_h$. Similar intuition underlies the third part of the Lemma.

Lemma 4 has two important implications. The first concerns which types can be pooling, and on which report.

Corollary 1. *If there is pooling on any report, it can only be on the highest report, and hence must include the highest type.*

Proof. Consider a pooled report $r_p < 1$. By Lemma 4, for small enough $\varepsilon > 0, \delta > 0$, a type $t_h(r_p) - \varepsilon$ would prefer to report $r_p + \delta$ and induce $a^R(t_h(r_p))$ rather than report r_p and induce $a < a^R(t_h(r_p))$, contradicting equilibrium. □

[21] As the proof makes clear, this must hold in any report monotone equilibrium, and does not require part 2 of the mD1 criterion.

Consequently, there can be at most one pool in an mD1 equilibrium. Since full separation has already been shown to be impossible, there must in fact be exactly one pool. The basic structure of reporting in any mD1 equilibrium is now clear: there must be some *cutoff* type $\hat{t} \in [0,1)$ such that there is full separation below $\hat{t}$, and pooling of all types above $\hat{t}$ on $r_p = 1$. Recall that this does not mean that all types above $\hat{t}$ are bunched, only that they all use the same report. The other implication of Lemma 4 is that if the the lowest type is separating, then it must be using the lowest report.

Corollary 2. *If type* 0 *is separating, then* $\rho(0) = 0$.

Proof. Suppose not, i.e. $\alpha(\rho(0), \mu(0)) = a^R(0)$ and $\rho(0) > 0$. By Lemma 4, for all m, $\alpha(0,m) = a^R(0)$. But then since $C(0,0) < C(\rho(0), 0)$, type 0 strictly prefers to play $(0,m)$ for any m, rather than $(\rho(0), \mu(0))$, contradicting equilibrium. □

Next, I characterize the cutoff type $\hat{t}$. A useful fact (Lemma A.4 in the Appendix) is that the function ρ is continuous everywhere except possibly at $\hat{t}$. Moreover, if $\hat{t} > 0$, then ρ is either left- or right-continuous at $\hat{t}$. This implies that if $\hat{t} > 0$, and $\lim_{t \uparrow \hat{t}} \rho(t) < 1$, then either $\hat{t}$ separates itself by reporting $\lim_{t \uparrow \hat{t}} \rho(t)$, or it pools with all higher types by reporting 1. On the other hand, when $\hat{t} = 0$, then by Corollary 2, $\rho(0) = 0$ if type 0 is separating, and if type 0 is not separating, then it must be that $\rho(0) = 1$. Equilibrium imposes an indifference condition for $\hat{t} > 0$, and a weak preference condition for $\hat{t} = 0$. To state these formally, let $m_1 \equiv \lim_{t \downarrow \hat{t}} \mu(t)$ and $r_1 \equiv \lim_{t \uparrow \hat{t}} \rho(t)$. Note that the former is well-defined by the fact that there is bunching at the bottom of the pool.

Lemma 5. *In any mD1 equilibrium, if the cutoff* $\hat{t} > 0$ *and* $r_1 < 1$ *then* $\hat{t}$ *is indifferent between playing* (r_1, m) *and playing* $(1, m_1)$*, for any message* m*. If* $\hat{t} = 0$*, type* 0 *weakly prefers playing* $(1, m_1)$ *to* $(0, m)$ *for any* m*.*

That the second statement of the Lemma must hold is clear. The rationale behind the first statement is the following. Suppose it were not true. Then type $\hat{t}$ strictly prefers playing $(1, m_1)$ to (r_1, m) for some message m. It is easy to check that for any message m, $\alpha(r_1, m) \geq a^R(\hat{t})$ by the fact that there is separation below $\hat{t}$ and report monotonicity. One can show that the mD1 criterion implies that in fact $\alpha(r_1, m) = a^R(\hat{t})$, simply because amongst all types $t \geq \hat{t}$, $\hat{t}$ has the most to gain by deviating from the pool down to r_1. So type $\hat{t}$ strictly prefers being perceived as the bottom bunch of the pool and incurring the cost of report 1, rather than being perceived as type $\hat{t}$ and incurring the cost of r_1. But then continuity implies that so will some type slightly lower than $\hat{t}$, contradicting separation below $\hat{t}$.

It remains to analyze the separating portion of the type space. Since ρ must be strictly increasing in this region, ρ^{-1} is well-defined, and optimality requires

$$\rho(t) \in \arg\max_r U\left(a^R\left(\rho^{-1}(r)\right), t\right) - kC(r,t)$$

Since ρ is increasing it is a.e. differentiable, and must satisfy the following first order condition at all points of differentiability:

$$U_1\left(a^R(t), t\right) a_1^R(t) \frac{1}{\rho'(t)} - kC_1(\rho(t), t) = 0$$

This is an ordinary non-linear differential equation. By Corollary 2, the initial condition is $\rho(0) = 0$, and the resulting initial value problem is therefore that specified by (DE).

Lemma 6. *There is a unique solution to (DE).*

The proof is complicated by the lack of a Lipschitz condition on $(t, \rho) \in [0, 1] \times [0, \infty)$. Existence is established by first considering a perturbed problem with initial condition of $\rho(0) = \varepsilon > 0$, proving existence of a solution to this problem (which still lacks a Lipschitz condition on the entire region, but does possess one in a neighborhood of the initial condition, unlike (DE)), then taking the limit of the solutions as $\varepsilon \to 0$, and showing that this defines a function which in fact solves (DE). Uniqueness is established by showing that any solution to (DE) must be the limit of solutions to the perturbed problems, and proving that that those are unique.

Let ρ^* denote the unique solution to (DE). The proof of the Lemma also shows that ρ^* is strictly increasing, continuously differentiable, and satisfies $\rho^*(t) \geq t$, with equality only for $t = 0$. Thus there is a unique solution to $\rho^*(t) = 1$, which I denote $\bar{t}$. It is easily verified that ρ^* is indeed a separating function below $\bar{t}$ in the sense that if the Receiver plays $(\rho^*)^{-1}(r)$ for all $r < 1$, then every type $t < \bar{t}$ prefers sending $\rho^*(t)$ to sending $\rho^*(t')$ for any $t' < \bar{t}$ (with the cheap talk message being irrelevant). It follows that $\bar{t} < 1$; otherwise there is a fully separating equilibrium, which cannot be the case by Proposition 1.

We can now conclude. Arguments similar to that of Mailath (1987) establish that in the separating part of the type space, the reporting strategy must in fact be exactly ρ^*. Thus, the previous discussion implies that any mD1 equilibrium involves separation below some type $\hat{t} \leq \bar{t}$ using the report strategy ρ^*, satisfies the indifference/weak preference condition for type $\hat{t}$, and has a partition of $\left[\hat{t}, 1\right]$ into some finite number of bunches using cheap talk. It is straightforward to see that each type on the boundary between two bunches on the pooling report must be indifferent between being perceived as a member of either bunch, just as in Crawford and Sobel (1982, Theorem 1). Obviously there can be an inessential multiplicity among equilibria in the sense that the cheap talk messages played in equilibrium are arbitrary, but this is of no economic consequence. This proves necessity of parts (i)-(iii) of the Theorem. Sufficiency is straightforward, since when the conditions are satisfied, the proposed strategies (a)-(c) constitute an mD1 equilibrium.

5.3 Existence

Thus far, I have only characterized necessary and sufficient conditions for an mD1 equilibrium. The next result assures that these conditions can always be met.

Theorem 2. *An mD1 equilibrium exists.*

The details of the proof are quite involved, but the basic idea is constructive, as follows. Assume k is small for the sake of illustration. Starting with type $\bar{t}$ and working down to type 0, traces the indifference curves for each type t in (a, r) space that keep type t indifferent with $\left(a^R(t), \rho^*(t)\right)$. These are elliptical and hence generally pass through the $r = 1$ line twice. For each type $t \leq \bar{t}$, there are two types (assume they exist) $p_1^l(t)$ and $p_1^r(t)$ such that he is indifferent between separating himself with report $\rho^*(t)$ and eliciting the responses $\bar{a}(t, p_1^l(t))$ or $\bar{a}(t, p_1^r(t))$ from the Receiver by sending report $r = 1$ and some cheap talk message. Thus if t is the candidate cutoff type for an mD1 equilibrium, $p_1^l(t)$ and $p_1^r(t)$ are the candidate boundary types in the first bunch on $r = 1$. The indifference conditions (A) then implies a unique sequence of $p_j^q(t)$ ($j = 2, ..., q \in \{l, r\}$) that define the successive boundaries of bunches on the report $r = 1$. An mD1 equilibrium exists when either (i) there is some type $t \in (0, \bar{t}]$ such that for some integer $j \geq 1$, $p_j^q(t) = 1$ for a $q \in \{l, r\}$; or (ii) there is an integer $j \geq 1$ such that $p_j^q(0) = 1$ for a $q \in \{l, r\}$. Continuity of each p_j^q is shown to guarantee that one of these two cases can be satisfied.

One consequence of the proof of Theorem 2, which I note here for completeness, is that when k is sufficiently large, there is always an equilibrium with a strictly positive measure of types at the bottom end of the type space that separate themselves.

Proposition 2. *There exists $\bar{k} > 0$ such that if $k > \bar{k}$, there is an mD1 equilibrium with a strictly positive measure of separating types.*

6 Almost-Cheap Talk

Given the existence result, I will henceforth refer to "mD1 equilibria" as just "equilibria" for brevity. The previous section established that any equilibrium for a given k is completely described by a cutoff $\hat{t}^k \geq 0$ and a sequence $\{t_0^k = \hat{t}^k, t_1^k, ..., t_{J-1}^k, t_J^k = 1\}$. I shall refer to such a sequence as an *equilibrium outcome*, or simply, an *outcome*.[22] We are now ready to consider the set of equilibria as k gets small, and compare them to the CS equilibria. First, I briefly recapitulate the relevant results of CS.

[22] This is a little non-standard since outcomes as I have defined them ignore the payoff-relevant misreporting cost for the Sender. However, this definition is simple and sufficient since the focus henceforth is on equilibria as $k \to 0$, hence payoffs are arbitrarily well approximated by U and the equilibrium partition.

6.1 CS Equilibria

The main result of CS (Crawford and Sobel 1982, Thereom 1) is that any equilibrium in the their model is essentially equivalent to one where the type space is partitioned into N segments $\{t_0 = 0, t_1, ..., t_{N-1}, t_N = 1\}$ such that each type only reveals which segment he is in, and upon hearing the message that the type is in (t_j, t_{j+1}), the Receiver plays $\overline{a}(t_j, t_{j+1})$. The boundaries of the segments must satisfy the arbitrage condition A for all $j \in \{1, \ldots, N-1\}$. Moreover, there exists an integer $\overline{N} \geq 1$ such that there is an equilibrium with N segments if and only if $N \in \{1, ..., \overline{N}\}$. The equilibrium with $N = 1$ is the *babbling equilibrium*, and this always exists. The magnitude of $\overline{N}$ depends upon the similarity of preference between S and R: the more similar they are, the higher is $\overline{N}$. Any equilibrium with $N > 1$ is said to be an *informative equilibrium.* CS also use the following condition for welfare and comparative statics (Crawford and Sobel 1982, p. 1444).

Condition M. *For any two increasing sequences,* $\{t_0, t_1, ..., t_K\}$ *and* $\{\tilde{t}_0, \tilde{t}_1, ..., t_K\}$ *that both satisfy the arbitrage condition* (A) *for* $j \in \{1, \ldots, K\}$*: if* $t_1 > \tilde{t}_1 > t_0 = \tilde{t}_0$*, then* $t_j > \tilde{t}_j$ *for all* $j \in \{1, \ldots, K\}$*.*

Although this is not stated in terms of primitives of the model, CS provide sufficient conditions on primitives that guarantee it (Crawford and Sobel 1982, Theorem 2). Roughly, Condition M requires that the preferred actions a^S and a^R shift similarly with t. All applied papers following CS use this assumption either implicitly (by choosing a specification that satisfies it) or explicitly. The reason is that Condition M guarantees certain attractive properties of equilibria, as summarized below. Let $t_j(N)$ denote the j^{th} boundary type ($j \in \{0, 1, \ldots, N\}$) in a N-segment CS equilibrium.

Lemma 7. ***(CS Lemma 3)*** *If Condition M holds, then*

1. *there is a unique CS partition equilibrium of size* $N \in \{1, \ldots, \overline{N}\}$*.*
2. *for all* $N \in \{1, \ldots, \overline{N} - 1\}$ *and* $j \in \{1, \ldots, N\}$*, ,* $t_j(N + 1) < t_j(N)$*.*

CS use this Lemma to prove that both players ex-ante (i.e. before S learns his type) strictly prefer equilibria with more segments,[23] which implies that the equilibrium with $\overline{N}$ segments is the ex-ante Pareto-dominant equilibrium; this is also referred to as the *most-informative* equilibrium. More generally, R's expected utility in any equilibrium is a measure of the informativeness of the equilibrium. Therefore, the Lemma says that under Condition M, equilibria are strictly ranked by informativeness, and an equilibrium with $N + 1$ segments is *more informative* than an equilibrium with N segments.

Remark 3. In what follows, Condition M is not assumed unless explicitly stated. However, it is important to emphasize that without Condition M, it is not guaranteed that (a) CS equilibria with more segments are more informative; nor (b) informativeness coincides with ex-ante Pareto dominance.

[23] Crawford and Sobel (1982, Theorems 3 and 5).

6.2 Small k Equilibria

In the ensuing discussion, I refer to any partition of the type space $[0,1]$ that is supported by a CS equilibrium as a *CS outcome* and denote it $\{t_0^0 = 0, t_1^0, \dots, t_N^0 = 1\}$. Recall that an outcome of the current model for a given k is described by a $\hat{t}^k \geq 0$ and $\{t_0^k = \hat{t}^k, t_1^k, ..., t_{J-1}^k, t_J^k = 1\}$. To maintain the distinction between CS and the current model, I shall often to refer to an equilibrium [outcome] of the current model as an "equilibrium [outcome] with reporting".

The analysis revolves around the following critical type. Define

$$t^* \equiv \begin{cases} t > 0 : U(\overline{a}(0,t),0) = U(a^R(0),0) \text{ if } U(\overline{a}(0,1),0) \leq U(a^R(0),0) \\ \infty \text{ otherwise} \end{cases}$$

That is, if $t^* \leq 1$, then $t^* > 0$ is the type such that the Sender of type 0 is indifferent between revealing his true type exactly and revealing only that he lies in in the non-degenerate interval $[0, t^*]$. If the type 0 Sender strictly prefers revealing no information at all to revealing his true type, then $t^* = \infty$. Since $\overline{a}$ is strictly increasing in both arguments so long as they are in the type space $[0,1]$ (and $\overline{a}$ is constant in the arguments outside the type space), and $\overline{a}(0,0) = a^R(0) < a^S(0)$, it follows that t^* is unique, and $t^* \in (0,1] \cup \{\infty\}$. Note that for any $t \in (0,1]$, $t < (>) t^* \Leftrightarrow U(\overline{a}(0,t),0) > (<) U(a^R(0),0)$.

I first confirm an analogue of "upper hemi-continuity", i.e. that every convergent sequence of outcomes with reporting converges to a CS outcome. The intuition stems from the observation that when k is small, the cutoff type, $\hat{t}^k$, in any equilibrium with reporting is close to 0, and the boundary conditions for segmentation within the reporting pool are the same as CS.

Proposition 3. *$\forall \epsilon > 0$, $\exists \delta > 0$ such that when $k < \delta$, for any equilibrium outcome with reporting, $\{\hat{t}^k = t_0^k, t_1^k, \dots, t_J^k = 1\}$, there is a CS outcome, $\{t_0^{CS} = 0, t_1^{CS}, \dots, t_N^{CS} = 1\}$, such that $N = J$ and $\left| t_j^k - t_j^{CS} \right| < \epsilon$ for all $j \in \{0, 1, \dots, J\}$.*

Proof. Fix an $\epsilon > 0$. From Theorem 1 and the definition of ρ^* (DE), it follows that for sufficiently small k, every equilibrium with reporting has $t_0^k < \epsilon$. It suffices to argue that t_1^k must be close to some t_1^0, because then the indifference conditions (A) in Theorem 1 assure that every boundary type in the equilibrium with reporting is close to the corresponding boundary type in the CS equilibrium. Suppose by way of contradiction that there is no t_1^0 such that $\left| t_1^k - t_1^0 \right| < \epsilon$. It cannot be that $t_0^k = 0$ because then there is a CS equilibrium with $t_1^0 = t_1^k$. Given that $t_0^k > 0$, Theorem 1 implies that

$$U(a^R(t_0^k), t_0^k) - U(\overline{a}(t_0^k, t_1^k), t_0^k) \quad = \quad k[C(\rho^*(t_0^k), t_0^k) - C(1, t_0^k)]$$

For k sufficiently small, the RHS is arbitrarily close to 0 (since C is bounded), and hence the LHS must be close to 0. Since t_0^k can be made arbitrarily close to 0 by picking k

small enough, it follows from continuity that $\left|t_1^k - t^*\right| < \epsilon$. By the hypothesis towards contradiction, there is no CS equilibrium with first segment $[0, t^*]$. Since there are only a finite number of CS equilibria, there is no CS equilibrium with first segment boundary in a small neighborhood of t^*. Equivalently, if $\{\tau_0 = 0, \tau_1 = t^*, \tau_2, \tau_3, \ldots\}$ is a solution to the the difference equation (A), there is a $\theta > 0$ such that no τ_j $(j \geq 0)$ lies in $(1-\theta, 1]$. Since each t_j^k in the equilibrium with reporting is arbitrarily close to τ_j when k is sufficiently small, it follows that there is no J such that $t_J^k = 1$. This however contradicts the conditions for an equilibrium with reporting. □

The more interesting issue is that of "lower hemi-continuity", or robustness of a given CS outcome. Two notions of robustness are useful, where implicitly this refers to robustness with respect to almost-cheap talk (i.e. the reporting dimension with k small).

Definition 3. A CS outcome, $\{t_0^0 = 0, t_1^0, \ldots, t_N^0 = 1\}$, is *robust* if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all $k < \delta$, there is an outcome with reporting $\{t_0^k, t_1^k, \ldots, t_N^k = 1\}$ where for all $j \in \{0, 1, \ldots, N\}$, $\left|t_j^k - t_j^0\right| < \epsilon$.

That is, a CS outcome is robust to almost-cheap talk if there is a sequence of outcomes with reporting that converges to it as $k \to 0$.

Definition 4. A CS outcome $\{t_0^0 = 0, t_1^0, \ldots, t_N^0 = 1\}$ is *strongly robust* if there exists $\delta > 0$ such that for all $k < \delta$, $\{0 = t_0^k = t_0^0, t_1^k = t_1^0, \ldots, t_N^k = t_N^0 = 1\}$ is an outcome with reporting.

In words, a CS outcome is strongly robust to almost-cheap talk if it remains an outcome when misreporting costs are sufficiently small. Plainly, strong robustness implies robustness.

It is useful to define $\kappa(t_1)$ as the cost that would make type 0 indifferent between inducing $a^R(0)$ with report 0, and pooling with report 1 given that the lowest bunch is defined by $[0, t_1]$. That is,

$$\kappa(t_1) \equiv \frac{U(\overline{a}(0, t_1), 0) - U\left(a^R(0), 0\right)}{C(1, 0) - C(0, 0)}$$

Note that $\kappa(t_1)$ may be non-positive, in which case type 0 is never indifferent between the relevant alternatives for any $k > 0$. Our first comparison result gives necessary and sufficient conditions for when a CS outcome can be an outcome of our model.

Proposition 4. *A CS outcome with first segment* $\left[0, t_1^0\right]$ *is an outcome with reporting if and only if* $k \leq \kappa\left(t_1^0\right)$.

Proof. (Necessity) If $k > \kappa\left(t_1^0\right)$, then by the definition of κ, type 0 strictly prefers to separate with report 0 rather than bunch with $\left[0, t_1^0\right]$ using report 1. But then, by Theorem 1 (iii), there cannot be an equilibrium outcome where the cutoff type is $\hat{t} = 0$ and the first bunch is $\left[0, t_1^0\right]$.

(Sufficiency) Since $k \leq \kappa\left(t_1^0\right)$, the CS outcome partition satisfies Theorem 1 with the cutoff type being $\hat{t}^k = 0$, hence it is an equilibrium outcome. □

Using this result, I now derive a simple condition that determines whether or not a CS outcome is robust.

Theorem 3. *A CS outcome with first segment $\left[0, t_1^0\right]$ is*

1. *strongly robust if and only if $t_1^0 < t^*$*
2. *not robust if $t_1^0 > t^*$*
3. *robust if $t_1^0 = t^*$ and Condition M holds*

Proof.

1. The first part is consequence of Proposition 4, since $t_1 < t^* \Leftrightarrow \kappa(t_1) > 0$. Hence if $t_1 \geq t^*$, then the CS outcome is not an outcome with reporting for any k, and if $t_1 < t^*$, it is an outcome for all $k < \kappa(t_1)$.

2. For the second part, suppose the statement is false. Then $t_1^0 > t^*$ (which implies that $t^* \in (0,1)$) and yet there is a sequence of equilibria with reporting such that for every $\delta > 0$, there exists an $\varepsilon > 0$ such that if $k < \varepsilon$ then the equilibrium for cost k has cutoff type (the highest type who is separated) $\hat{t}^k < \delta$ and the boundary of the first bunch, call it t_1^k, satisfies $\left|t_1^k - t_1^0\right| < \delta$. I now argue to a contradiction. Since $t_1^0 > t^*$, $U(a^R(0), 0) > U(\overline{a}(0, t_1^0), 0)$. For small enough k, $\hat{t}^k$ and t_1^k are arbitrarily close to 0 and t_1^0 respectively, and hence by continuity of $\overline{a}$ and U, for small enough k,

$$U\left(a^R\left(\hat{t}^k\right), \hat{t}^k\right) > U\left(\overline{a}\left(\hat{t}^k, t_1^k\right), \hat{t}^k\right)$$

Since $C\left(1, \hat{t}^k\right) \geq C\left(\rho^*\left(\hat{t}^k\right), \hat{t}^k\right)$ (because $1 \geq \rho^*\left(\hat{t}^k\right) \geq \hat{t}^k$), this implies that for small enough k,

$$U\left(a^R\left(\hat{t}^k\right), \hat{t}^k\right) - kC\left(\rho^*\left(\hat{t}^k\right), \hat{t}^k\right) > U\left(\overline{a}\left(\hat{t}^k, t_1^k\right), \hat{t}^k\right) - kC\left(1, \hat{t}^k\right)$$

which means that type $\hat{t}^k$ strictly prefers separating to bunching with $\left[\hat{t}^k, t_1^k\right]$, contradicting equilibrium (Theorem 1 (ii,iii)).

3. The third part is deferred to the Appendix. □

Fix a CS equilibrium partition with first segment $[0, t_1]$. The condition that $t_1 < t^*$ requires that the lowest type strictly prefers the action induced by pooling with the lowest segment of the partition over the action he would induce if he separated himself. It is crucial to emphasize that $t_1 < t^*$ is *not* part of the equilibrium requirement in CS. CS equilibrium only requires that given a partition, each type — and in particular, the lowest type — weakly prefer the action induced by truthfully revealing which segment he is in over claiming that he is in some other segment of the partition. In equilibrium, a type does not have the ability to separate himself even if he would desire to. However, in the current model with reporting, any type below the cutoff type *does* have the ability to separate itself in a given equilibrium. Hence, in *any* equilibrium, the lowest type can separate itself. It is this feature that places a restriction on the set of CS equilibria that are robust to almost-cheap talk.

Turning to the second part of the Theorem, it does not just say that when $t_1 > t^*$, the relevant CS outcome is not an outcome with reporting for any k (this fact is immediate from Proposition 4), but moreover, that no sequence of equilibrium outcomes converges to the CS outcome. So when $t_1 > t^*$, it is not the case that with almost-cheap talk, there is an equilibrium outcome that "approximates" the CS outcome. On the other hand, when $t_1 < t^*$, the first part of the Theorem says the CS outcome itself is an equilibrium outcome with almost-cheap talk, and hence in a sense extremely robust to the introduction of small misreporting costs. The third part of the Theorem says that at least when Condition M holds, a CS outcome with first segment boundary t^* is the limit of equilibrium outcomes with reporting.[24]

The issue of robustness therefore reduces to which CS equilibrium outcomes satisfy $t_1 < t^*$, and which do not. Clearly, if some CS outcome is strongly robust, then so is every CS outcome with shorter first segment. Conversely, if some CS outcome is not robust, then neither is any CS outcome with longer first segment. To facilitate exposition, a simple observation about CS equilibria is useful. By 'generic' below, I mean for "almost all" preferences that satisfy the assumptions of the model.[25,26]

Observation 2. *For generic preferences, no CS equilibrium has a first segment* $[0, t^*]$.

This is intuitive, since given U and V, if the first segment boundary $t_1 = t^*$, each successive segment boundary t_j ($j \geq 2$) is uniquely pinned down by the arbitrage condition

[24] While it is natural to conjecture that this generalizes even without Condition M, a proof has proved elusive. Proposition 5 extends to cover the case when there is no CS outcome with first segment boundary smaller than t^*. In any case, this issue is not especially important, since cases where there is a CS outcome with first segment boundary t^* are non-generic, as noted in Observation 2.

[25] Formally, let $\mathcal{U}$ be the set of all (U, V) pairs that satisfy the assumptions of the model. Every pair $(U, V) \in \mathcal{U}$ induces a well-defined pair (a^S, a^R). Let $\mathcal{A}$ be the space of all such (a^S, a^R) pairs. Endow $\mathcal{U}$ with the topology induced by the product topology of uniform $\mathcal{C}^1$ convergence on $\mathcal{A}$ (so that two pairs (U, V) and $(\tilde{U}, \tilde{V})$ are close if their induced pairs (a^S, a^R) and $(\tilde{a}^S, \tilde{a}^R)$ have values and first derivatives that are close everywhere on $[0, 1] \times [0, 1]$). A property is said to hold generically on $\mathcal{U}$ if it holds for an open-dense subset of $\mathcal{U}$.

[26] The definition of genericity given above is topological, and this is convenient given the difficulties with measure-theoretic notions of "almost all" in infinite dimensional subspaces. However, following ideas in Anderson and Zame (2000), a measure-theoretic definition can be provided.

(A), and this sequence will form a CS equilibrium if and only if there is some $J \geq 1$ such that $t_J = 1$, which will generally not be true. The reason to make this observation is that it implies that for generic preferences, *every* CS equilibrium satisfies either $t_1 < t^*$ or $t_1 > t^*$, and by Theorem 3, is either strongly robust or not robust. Now I turn to identifying which of the CS equilibria satisfies which of the two inequalities.

Lemma 8. *At least one CS outcome has first segment $\left[0, t_1^0\right]$ satisfying $t_1^0 \leq t^*$. Moreover, if Condition M hold, there there is only such CS outcome.*

This is trivial if $t^* \geq 1$, so suppose that $t^* < 1$. The proof is constructive. Briefly, start by considering the type t^* such that type 0 is indifferent between actions $a^R(0)$ and $\overline{a}(0, t^*)$. For every type $t \in [0, t^*]$, set $p_0(t) = 0$ and $p_1(t) = t$. Successively construct the segment boundaries $p_j(t)$ ($j \geq 2$) that would keep type $p_{j-1}(t)$ indifferent between $\overline{a}(p_{j-2}(t), p_{j-1}(t))$ and $\overline{a}(p_{j-1}(t), p_j(t))$. The key is to observe that for all $j \geq 1$, $p_j(0) = p_{j-1}(t^*)$ by definition. There will be some (unique) integer $M \geq 1$ such that $p_M(0) = p_{M-1}(t^*) < 1 \leq p_M(t^*)$, and by continuity of p_M, some t' such that $p_M(t') = 1$. By construction, the partition $\{0 = p_0(t'), t' = p_1(t'), p_2(t'), \ldots, p_M(t') = 1\}$ is a CS equilibrium with first segment boundary $t' \leq t^*$.

The result says that there is always at least one CS equilibrium in which the lowest type weakly prefers the action he induces by pooling with the lowest segment over the action he would induce *if* he could separate himself. In particular, the CS equilibrium with shortest first segment satisfies this. The second part of the results says that under Condition M, only the most-informative equilibrium (the one with the shortest first segment) satisfies this property. By the earlier observation, it follows that generically, this preference will in fact be strict. When combined with Theorem 3, this yields an important conclusion.

Proposition 5. *At least one CS outcome is robust, and generically, at least one CS outcome is strongly robust.*

Proof. The second part is an immediate consequence of the preceding discussion. To prove the first part, I argue that the CS outcome with smallest first segment amongst all CS outcomes is robust. Denote this segment $[0, t_1^0]$. By Lemma 8, $t_1^0 \leq t^*$. If $t_1^0 < t^*$, then Theorem 3 implies that the CS equilibrium is strongly robust and hence robust, and we are done. So assume $t_1^0 = t^*$; hence in every CS outcome type 0 gets no more utility than he would through separation. Pick any sequence of outcomes with reporting indexed by k, $\{t_0^k = \hat{t}^k, t_1^k, \ldots, t_{J(k)}^k = 1\}$. It must be that for all k, $\hat{t}^k > 0$. This is because any outcome with reporting with cutoff of type 0 is a CS outcome, and due to the presence of the misreporting cost, type 0 *strictly* prefers separation to being pooled with the first segment. But by hypothesis, there is no such CS outcome. By Theorem 1 (ii), equilibrium with reporting requires

$$U(\overline{a}(\hat{t}^k, t_1^k), \hat{t}^k) - U(a^R(\hat{t}^k), \hat{t}^k) = k[C(1, \hat{t}^k) - C(\rho^*(\hat{t}^k), \hat{t}^k)]$$

Note that the RHS is weakly positive for all $k > 0$. Hence the LHS is weakly positive. It is easy to check that $\bar{t}$ is strictly decreasing as k decreases, hence $\hat{t}^k$ can be made arbitrarily small by choosing k small enough. It is now sufficient to prove that by picking k small enough, we can also make t_1^k arbitrarily close to t_1^0, since then all subsequent boundary types t_j^k $(j \geq 2)$ will also be arbitrarily close to t_j^0. To see that we can do this, observe that since $C(1,t) - C(\rho^*(t), t)$ is bounded for all $t < \bar{t}$, the RHS above is approaching 0 as $k \to 0$. Hence the LHS must also approach 0. Since $\hat{t}^k$ is arbitrarily close to 0 for small enough k, this requires that t_1^k be arbitrarily close to t^*, and by hypothesis $t^* = t_1^0$. □

Hence, there is always a CS outcome that is the limit of equilibrium outcomes with reporting, and moreover, generically, at least one CS outcome remains an outcome with reporting once k is sufficiently small. This assures that if one interprets robustness to almost-cheap talk as a selection criterion amongst CS equilibria, then at least once CS equilibrium survives the criterion. The next result gives a partial converse: it says that under Condition M, *only one* equilibrium survives.

Theorem 4. *Assume Condition M. Only the most-informative CS outcome is robust, and generically, it is strongly robust.*

Proof. Straightforward consequence of Theorem 3, Lemma 8, and Observation 2. □

Thus, under the standard regularity assumption of Condition M, there is a unique CS outcome that is the limit of outcomes with reporting: the most-informative CS outcome. It is worth emphasizing the following implication.

Corollary 3. *Assume Condition M. The babbling outcome is not robust unless it is the unique CS outcome.*

Remark 4. Recall that even when Condition M fails, if a CS outcome with first segment $[0, t_1]$ is an outcome with almost-cheap talk, then so is a CS outcome with first segment $[0, \tilde{t}_1]$, for any $\tilde{t}_1 \leq t_1$ (Theorem 3). This implies that if the babbling outcome *is* robust to almost-cheap talk, then so is every other CS outcome. Condition M guarantees that there is no other CS outcome in this case.

I end this section by noting that the leading special case of the CS model, the "uniform-quadratic" setup does satisfy Condition M. Numerous papers focus on this case due to its analytic tractability.[27] The above results can be viewed as justifying the selection of the most-informative equilibrium for this setting.

[27] See for example Crawford and Sobel (1982, Section 4), de Garidel-Thoron and Ottaviani (2000), Gilligan and Krehbiel (1987), Harris and Raviv (2002), Krishna and Morgan (2004), and Li (2004) among others.

7 Discussion

7.1 Relation to Other Refinements

Arguably the most well-known cheap talk refinement in the literature is the concept of *neologism-proof equilibrium* (hereafter, NPE), due to Farrell (1993). The criterion is follows: assume that given any equilibrium of a cheap talk game, for every subset of types, $\Omega \subseteq T$, there exists a distinct neologism, or out-of-equilibrium message, m_Ω. For any $\Omega \subseteq T$, define $r(\Omega)$ as the optimal response for the Receiver if the only information she has is that $t \in \Omega$. An equilibrium is *not* a NPE if there exists some set $\Omega \subseteq T$ s.t. a type t strictly prefers action $r(\Omega)$ to the equilibrium action he induces if and only if $t \in \Omega$.

The NPE concept is developed for general cheap talk games, and unfortunately its applicability to CS is limited. In the CS model, the set of NPE is generally empty whenever babbling is not the unique outcome (Farrell 1993, p.529).[28] In contrast, the approach in this paper guarantees that at least one CS equilibrium survives, while simultaneously ruling out some CS equilibria in general. On the other hand, it is not clear that the present approach will generalize to arbitrary cheap talk games, whereas NPE can be a useful tool in those settings where it exists.[29]

That being said, it is useful to connect the two approaches. In the current model, observe that when k is small, all reports but the highest one are neologisms in any equilibrium with reporting that supports a CS outcome. However, and this is the crucial point, because these reports are *not* cheap talk, I have appealed to out-of-equilibrium reasoning to determine their equilibrium meaning.[30] A few sentences from Farrell (1993, p. 519) may be instructive:

> "[First, the] meaning of messages that are used in equilibrium ... is established by Bayes' rule, which tells us their meaning-in-use. Secondly, ... [a message] may have a meaning that can be determined, or at least somewhat restricted, by introspection. This yields restrictions on out-of-equilibrium beliefs in generic signaling games: but they do not apply to cheap-talk games. Finally ... a message may have a focal meaning, if it is phrased in a preexisting language."

On the costly reporting dimension, regardless of k, we ultimately only have messages of the first and second kind, not of the third. More precisely, it is not that the messages *must not* have a focal meaning, but instead that the first and second considerations mentioned above override any others, including focal meanings, and pin down the meaning of all reports

[28]For the same reasons, none of the CS equilibria are *announcement proof* in the sense of Matthews, Okuno-Fujiwara, and Postlewaite (1991).

[29]Grossman and Perry (1986) define *perfect sequential equilibrium* for general signaling games. This concept is in the spirit of NPE, and can also suffer from non-existence. Grossman and Perry (1986, p. 110) suggest that it [only] be used whenever it exists.

[30]Of course, as always, this is "endogenous" with respect to the assumptions of the model.

on the costly dimension. In particular, all out-of-equilibrium reports mean "I am the lowest type" and are responded to accordingly by the Receiver. In this sense, all neologisms in the relevant equilibria have *the same meaning*, despite the existence of a plethora of unused reports.[31]

The critical difference between Farrell (1993) and this paper is that I start out by specifying completely the set of available messages. Thus, each of these messages has a meaning that is determined based upon usage or introspection, whereas his approach relies upon messages whose meaning is determined outside the model. Consequently, *which* messages are neologisms with respect to any equilibrium (in addition to what they *mean*) is determined within the model in this paper, unlike in Farrell (1993). So while both approaches share the common feature of augmenting the CS model with "extra" messages, how these messages influence a putative CS outcome is very different.[32]

7.2 Extensions

7.2.1 Other Report Spaces

The analysis thus far assumed that the Sender must sender a report $r \in T$. While I suggested that this is the natural assumption for many settings, one might imagine that the report space could be different. So suppose instead r must lie in some set $\mathcal{R} \equiv [0, P]$, with $P \in (0, \infty]$, with some abuse of notation.

Theorem 5. *There exists $k^*(P) \in [0, \infty]$ such that a fully separating equilibrium exists if and only if $k \geq k^*(P)$. Moreover,*

1. *k^* is weakly decreasing in P and strictly decreasing for $P \geq 1$*
2. *$k^*(1) = \infty$*

[31]On a related note, notice that the neologisms in our equilibrium for small k are *not* credible in the sense that Farrell defines credibility of a neologism. Even though the Receiver believes that the Sender is of the lowest type when she sees any neologism, it is not the case that the only type of Sender who prefers $a^R(0)$ to the response he induces in equilibrium is the Sender of type 0. Recall that generically in those CS equilibrium that fail our selection criterion, $U(\overline{a}(0, t_1), 0) < U(a^R(0), 0)$. By continuity, all types ϵ small enough prefer getting $a^R(0)$ to what they get in equilibrium. So in the sense of Farrell, none of the neologisms available in our model are credible. This point is not really a puzzle though: the notion of credibility that Farrell defines for a neologism is appropriate for a neologism whose meaning is given *outside* the model - the third source of meaning in the earlier quote. For those neologisms that derive meaning through introspection within the model (as is the case with all our neologisms, at the risk of belaboring the point), their credibility and the Receiver's beliefs upon seeing them are determined accordingly.

[32]To emphasize the point, one can augment the current model to include the out-of-model neologisms of Farrell (1993), and carry over the definition of neologism-proofness appropriately. Based upon the logic of his non-existence result for the CS model, it is clear that the same problem would apply to our model. As an example, consider the uniform-quadratic setup (with quadratic misreporting costs), with $b = .2$. Then when $k = .02$, it can be checked that there is a unique mD1 equilibrium, and this involves complete pooling on the costly report, and implementation of the CS 2-segment partition {0,.1,1} through cheap talk. But the same neologism that breaks this equilibrium for CS - "I belong to the set $[.43333, 1]$" - breaks the mD1 equilibrium.

3. $k^*(\infty) = 0$

4. $k^*(P) > 0$ *for all* $P < \infty$

The theorem says that regardless of how large the report space is, as long as it is finite, there cannot be full separation in equilibrium once misreporting costs are small enough. The no full separation result in Proposition 1 earlier is a special case of this Theorem (Part 2 above). Note that if the report space is unbounded, then there will be a fully separating equilibrium for any $k > 0$.

7.2.2 Cost Functions

It was thus far assumed that the the most-preferred report for each type is the truth, i.e. that $C_1(t,t) = 0$ for all $t \in T$. As already noted, I consider this to be the most natural assumption given the interpretation of C as misreporting costs. However, a close examination of all the arguments will reveal that nothing substantive changes if instead there is some strictly increasing function $\theta : T \to T$ such that $C_1(\theta(t), t) = 0$ for all t.[33] So what is important is that the most-preferred report is monotonic in type. This extension can be useful in allowing for more general interpretations of the cost function. For example, as I discuss in more detail in Remark B.1 in the Appendix, this permits costs to stem from "crazy" types of Receivers, as introduced by Ottaviani and Squintani (2004).

7.2.3 Limited Costly Talk

In the main part of the paper, I discussed robustness of a cheap talk equilibrium to the availability of costly signaling as the magnitude of costs shrinks, i.e. $k \to 0$. There is another alternative that might be conceptually appealing in certain applied cases. Consider the report space being $[0, P]$. Instead of taking $k \to 0$, one could instead fix the magnitude k, and then take the limit $P \to 0$. Clearly, the case of $P = 0$ is equivalent to the CS model. It is straightforward given the main arguments of the paper that the condition for robustness of a CS equilibrium does *not* change whether one studies $k \to 0$ or $P \to 0$.

7.3 Other Equilibria

Thus far, I have focussed on equilibria that satisfy the mD1 criterion. I now discuss what happens if this is criterion is dropped. Without imposing monotonicity (of Sender's reporting strategy and Receiver's beliefs), it is not clear how much can be said about robustness of CS equilibria in general. I illustrate by proving that in the uniform-quadratic model, babbling can be supported as a D1 equilibrium once misreporting costs are sufficiently small.

Proposition 6. *In the U-Q model, babbling is a D1 equilibrium for small enough* k.

[33] cf. Bernheim and Severinov (2003).

In the U-Q model, the action elicited in equilibrium is $\overline{a}(0,1) = \frac{1}{2}$, whereas $a^R(0) = 0$. If $U(\frac{1}{2}, 0) > U(0, 0)$ then the proposition is immediate since Theorem 3 assures that babbling can be supported for small enough k as an mD1 equilibrium, hence a D1 equilibrium. So assume $U(\frac{1}{2}, 0) \leq U(0, 0)$, or equivalently the bias parameter b is no greater than $\frac{1}{4}$. The proposition is proved by showing that all types pooling on $r = \frac{1}{2} - b$ is part of a D1 equilibrium once k is sufficiently small.

Another possibility is to require monotonicity but drop the further belief restriction of D1. It remains an open question whether there is sequence of monotonic equilibria to this model that converges to a given CS equilibrium. The difficulty is characterizing the set of monotonic equilibria (even when k is small).

8 Conclusion

This paper has presented a model of communication between an informed Sender and an uninformed Receiver, wherein the Sender uses two kinds of messages: a pure costless message, and message where misreporting private information is costly. In the absence of the costly message, the model reduces to that of CS. Using a refinement approach on the costly message dimension, I have analyzed a class of equilibria that are appealing. By analyzing these equilibria when the magnitude of costly misreporting shrinks to arbitrarily small amounts, a theory of equilibrium selection amongst CS equilibria emerges. The selection criterion of robustness to almost-cheap talk is summarized by a very simple condition: given a CS equilibrium, one only needs to check whether lowest type prefers being pooled with the lowest segment of types to separating himself. At least one CS equilibrium satisfies this property, and under a regularity condition (Condition M), the most-informative equilibrium survives whereas the least-informative equilibrium is eliminated so long as there is more than one CS equilibrium. While most of the paper focused on the equilibrium selection aspect of the analysis, I have also suggested that the equilibrium characteristics with almost-cheap talk are important in their own right.

A potentially interesting issue for future research to pursue the approach of this paper to other classes of cheap talk games, for example games where more than one player possesses and communicates private information before actions are taken.

Appendix A: Proofs

Proof of Lemma 1 on Page 8. Suppose not, by way of contradiction. Then there exists $\hat{t} \in \left(0, \tilde{t}\right]$ s.t. $\rho\left(\hat{t}\right) = \hat{t}$. For small $\varepsilon \geq 0$, define $g(\varepsilon)$ as the expected utility gain for a type $\hat{t} - \varepsilon$ by deviating from $\rho\left(\hat{t} - \varepsilon\right)$ to $\rho\left(\hat{t}\right)$. Since by hypothesis all types below and including $\hat{t}$ are separating, we have

$$g(\varepsilon) \equiv \left[U\left(a^R\left(\hat{t}\right), \hat{t} - \varepsilon\right) - kC\left(\hat{t}, \hat{t} - \varepsilon\right)\right] - \left[U\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right) - kC\left(\rho\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right)\right]$$

That $C\left(\rho\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right) \geq C\left(\hat{t} - \varepsilon, \hat{t} - \varepsilon\right)$ implies

$$g(\varepsilon) \geq \phi(\varepsilon) \equiv \left[U\left(a^R\left(\hat{t}\right), \hat{t} - \varepsilon\right) - kC\left(\hat{t}, \hat{t} - \varepsilon\right)\right] - \left[U\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right) - kC\left(\hat{t} - \varepsilon, \hat{t} - \varepsilon\right)\right]$$

Clearly $\phi(0) = 0$. Differentiating yields

$$\begin{aligned} \phi'(\varepsilon) &= -U_2\left(a^R\left(\hat{t}\right), \hat{t} - \varepsilon\right) + kC_2\left(\hat{t}, \hat{t} - \varepsilon\right) - kC_1\left(\hat{t} - \varepsilon, \hat{t} - \varepsilon\right) - kC_2\left(\hat{t} - \varepsilon, \hat{t} - \varepsilon\right) \\ &\quad + U_1\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right) x_1^R\left(\hat{t} - \varepsilon\right) + U_2\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t} - \varepsilon\right) \end{aligned}$$

Recalling that $C_1\left(\hat{t}, \hat{t}\right) = 0$, it follows that $\phi'(0) = U_1\left(a^R\left(\hat{t}\right), \hat{t}\right) x_1^R\left(\hat{t}\right) > 0$. So for sufficiently small $\varepsilon > 0$, $g(\varepsilon) \geq \phi(\varepsilon) > 0$, implying that a type $\hat{t} - \varepsilon$ strictly prefers to imitate $\hat{t}$, contradicting equilibrium separation. □

Proof of Proposition 1 on Page 8. Assume that there is a separating equilibrium. By Lemma 1, $\rho(t) \neq t$ for all $t > 0$ (implying in particular that $\rho(1) < 1$) . The next step is to establish that $\rho(0) = 0$ and $\rho(t) < t$ for all $t < 1$. To do this, it suffices to show that if $\rho\left(\bar{t}\right) > \bar{t}$ for some $\bar{t} \in [0, 1)$, then $\rho(t) \geq t$ for all $t \in \left[\bar{t}, 1\right]$, since we already know that $\rho(1) < 1$.

<u>Claim</u>: If $\rho\left(\bar{t}\right) > \bar{t}$ for some $\bar{t} \in [0, 1)$, then $\rho(t) \geq t$ for all $t \in \left[\bar{t}, 1\right]$.

<u>Proof</u>: Suppose not. Then there exists $\bar{t} \in [0, 1)$ such that $\rho\left(\bar{t}\right) > \bar{t}$, and some $t' \in \left(\bar{t}, 1\right]$ such that $\rho(t') < t'$. Define

$$t^s = \sup_{t \in \left[\bar{t}, t'\right]} \{t : \rho(t) > t\}$$

By Lemma 1, there are two possibilities: (a) $\rho(t^s) < t^s$; (b) $\rho(t^s) > t^s$.[34] Suppose first $\rho(t^s) < t^s$. Then there exists an increasing sequence $t_n \nearrow t^s$ with $\rho(t_n) > t_n$. Let the limit of $\rho(t_n)$ be $\overline{\rho}$ (if necessary, take a convergent subsequence, whose existence is assured since this is a bounded sequence.) Obviously, $\overline{\rho} \geq t^s > \rho(t^s)$. Equilibrium incentive compatibility implies that for all n,

$$\begin{aligned} U\left(a^R(t_n), t_n\right) - kC\left(\rho(t_n), t_n\right) &\geq U\left(a^R(t^s), t_n\right) - kC\left(\rho(t^s), t_n\right) \\ U\left(a^R(t^s), t^s\right) - kC\left(\rho(t^s), t^s\right) &\geq U\left(a^R(t_n), t^s\right) - kC\left(\rho(t_n), t^s\right) \end{aligned}$$

Rearranging and adding yields

$$\begin{aligned} & U\left(a^R(t^s), t^s\right) - U\left(a^R(t_n), t^s\right) - \left[U\left(a^R(t^s), t_n\right) - U\left(a^R(t_n), t_n\right)\right] \\ \geq \; & k\left[C\left(\rho(t^s), t^s\right) - C\left(\rho(t_n), t^s\right) - \left(C\left(\rho(t^s), t_n\right) - C\left(\rho(t_n), t_n\right)\right)\right] \end{aligned}$$

[34] More explicitly, if $t^s > 0$, then $\rho(t^s) \neq t^s$ from Lemma 1. If $t^s = 0$, then it must be that $t^s = \bar{t} = 0$ and by hypothesis, $\rho\left(\bar{t}\right) > \bar{t}$.

This can be rewritten as

$$\int_{t_n}^{t^s}\int_{a^R(t_n)}^{a^R(t^s)} U_{12}(a,b)\,dadb \geq k\int_{t_n}^{t^s}\int_{\rho(t_n)}^{\rho(t^s)} C_{12}(a,b)\,dadb$$

Since U and C are smooth, $U_{12}(a,b)$ is bounded above by some constant $\upsilon > 0$ for any a, b, and similarly C_{12} is bounded below by some $\eta < 0$. So we have that for all n,

$$(t^s - t_n)\left(a^R(t^s) - a^R(t_n)\right)\upsilon \geq k(t^s - t_n)(\rho(t^s) - \rho(t_n))\,\eta$$

or equivalently

$$\left(a^R(t^s) - a^R(t_n)\right)\upsilon \geq k(\rho(t^n) - \rho(t_s))\,|\eta|$$

But the LHS is converging down to 0,whereas the RHS is converging up to $k(\overline{\rho} - \rho(t^s))\,|\eta| > 0$, contradicting the inequality for large enough n.

Now consider case (b), $\rho(t^s) > t^s$. Note that this implies $t^s < 1$. So there is a decreasing sequence $t_n \searrow t^s$ with $\rho(t_n) < t_n$.[35] The rest of the argument proceeds mutatis mutandis to case (a). Let the limit (of a convergent subsequence if necessary) of $\rho(t_n)$ be $\overline{\rho}$. Obviously, $\overline{\rho} \leq t^s < \rho(t^s)$. Equilibrium incentive compatibility implies that for all n,

$$\begin{aligned} U\left(a^R(t_n), t_n\right) - kC(\rho(t_n), t_n) &\geq U\left(a^R(t^s), t_n\right) - kC(\rho(t^s), t_n) \\ U\left(a^R(t^s), t^s\right) - kC(\rho(t^s), t^s) &\geq U\left(a^R(t_n), t^s\right) - kC(\rho(t_n), t^s) \end{aligned}$$

Rearranging, adding, and rewriting yields

$$\int_{t^s}^{t_n}\int_{a^R(t^s)}^{a^R(t_n)} U_{12}(a,b)\,dadb \geq k\int_{t^s}^{t_n}\int_{\rho(t^s)}^{\rho(t_n)} C_{12}(a,b)\,dadb$$

Defining $\upsilon > 0$ and $\eta < 0$ as before, we have that for all n,

$$\left(a^R(t_n) - a^R(t^s)\right)\upsilon \geq k(\rho(t_n) - \rho(t^s))\,\eta$$

But the LHS is converging down to 0, whereas the RHS is converging up to $k(\overline{\rho} - \rho(t^s))\,\eta > 0$, contradicting the inequality for large enough n. $\|$

Combined with Lemma 1, we have established that $\rho(0) = 0$ and $\rho(t) < t$ for all $t > 0$. Now we argue that ρ must be continuous.

Claim: ρ is continuous.

Proof: By way of contradiction, suppose there is a discontinuity at some t'. Take a sequence $t_n \to t'$, and let $\overline{\rho}$ be the limit of (if necessary, a convergent subsequence of) $\rho(t_n)$.[36] There are two cases: either $\overline{\rho} < \rho(t')$, or $\overline{\rho} > \rho(t')$.

Consider first $\overline{\rho} < \rho(t')$. Note that this requires $t' \neq 0$. So we have $t' > \rho(t') > \overline{\rho}$. Since $C(\cdot,\cdot)$ is continuous, there exists $\varepsilon > 0$ and N such that for all $n > N$, $C(\rho(t_n), t_n) - C(\rho(t'), t_n) > \varepsilon$. Moreover, by picking n large enough, we can make $U\left(a^R(t_n), t_n\right) - U\left(a^R(t'), t_n\right)$ arbitrarily close to 0. It follows that for large enough n, t_n prefers to imitate t', contradicting equilibrium.

Consider next $\overline{\rho} > \rho(t')$. Note that $\overline{\rho} < t'$, because otherwise for sufficiently large t_n, we would have $\rho(t_n) > t_n$. So we have $t' > \overline{\rho} > \rho(t')$. Since $C(\cdot,\cdot)$ is continuous, there exists $\varepsilon > 0$

[35] We know this because since $\rho(t') < t'$, it must cannot be that $t^s = t'$, and hence $t^s < t'$, and by definition of t^s and step 1, $\rho(t) < t$ for all $t \in (t^s, t']$.

[36] Since ρ is bounded, there is a convergent subsequence.

and N such that for all $n > N$, $C(\rho(t'), t') - C(\rho(t_n), t') > \varepsilon$. Moreover, by picking n large enough, we can make $U\left(a^R(t_n), t'\right) - U\left(a^R(t'), t'\right)$ arbitrarily close to 0. It follows that for large enough t, t' prefers to imitate t_n, contradicting equilibrium. $\|$

We conclude as follows. Since ρ is continuous with $\rho(0) = 0$, it must be weakly increasing on $(0, t')$ for some $t' > 0$. Given that $\rho(t) < t$ for all $t > 0$, we can choose $\varepsilon > 0$ small enough to make $U\left(a^R(t'), t' - \varepsilon\right) > U\left(a^R(t' - \varepsilon), t' - \varepsilon\right)$ and yet $C(\rho(t'), t' - \varepsilon) \leq C(\rho(t' - \varepsilon), t' - \varepsilon)$. But then type $t' - \varepsilon$ prefers to imitate type t', contradiction. □

Proof of Lemma 3 on Page 14. Given that in a report-monotone equilibrium, the set of types using any report is connected, the argument is completely analogous to Crawford and Sobel (1982, Theorem 1), hence omitted. □

Lemma A.1. *If in a report-monotone equilibrium there is pooling on a report $r_p < 1$, then there exists $\theta > 0$ such that reports $r \in (r_p, r_p + \theta)$ are not sent in equilibrium.*

Proof. Suppose there is pooling on $r_p < 1$. Define $t_h \equiv \sup\{t : \rho(t) = r_p\}$. If $\rho(t_h) > r_p$, then by report monotonicity, we are done, since reports in $(r_p, \rho(t_h))$ are unused. So assume that $\rho(t_h) = r_p$. Similarly, if $t_h = 1$, then we are done, since reports $r \in (r_p, 1)$ are unused. So assume $t_h < 1$. Let $m_h \equiv \mu(t_h)$.

Claim: There must be non-trivial bunching on (r_p, m_h).

Proof: If not, then since bunches are connected, there is no other type $t < t_h$ playing (r, m_h), whence type t_h is separating. But then, for small enough $\varepsilon > 0$, some type $t_h - \varepsilon$ would prefer to mimic type t_h, contradicting equilibrium. $\|$

So there is bunching on (r_p, m_h), whence $\alpha(r_p, m_h) < a^R(t_h)$. Let $\rho' \equiv \lim_{t \downarrow t_h} \rho(t)$. ($\rho'$ is well-defined by report monotonicity, though it may not be played in equilibrium.)

Claim: $\rho' > r_p$.

Proof: Suppose not. By report monotonicity, it must be that $\rho' = r_p = \rho(t_h)$. Note that ρ is then continuous at t_h. Since $\rho(t) > r_p$ for all $t > t_h$, it follows that ρ is strictly increasing on $(t_h, t_h + \delta)$ for some $\delta > 0$. Hence, defining $a_\varepsilon \equiv \alpha(\rho(t_h + \varepsilon), \mu(t_h + \varepsilon))$, we have $a_\varepsilon = a^R(t_h + \varepsilon)$ for small enough $\varepsilon > 0$. By picking $\varepsilon > 0$ small enough, we can make $C(\rho(t_h + \varepsilon), t_h) - C(r_p, t_h)$ arbitrarily close to 0, whereas $U(a_\varepsilon, t_h) - U(\alpha(r_p, m_h), t_h)$ is positive and bounded away from 0, because $\alpha(r_p, m_h) < a^R(t_h) < a_\varepsilon < a^S(t_h)$. Therefore, for small enough $\varepsilon > 0$, t_h prefers to imitate $t_h + \varepsilon$, contradicting equilibrium. $\|$

This completes the argument because reports in (r_p, ρ') are unused. □

Lemma A.2. *In any mD1 equilibrium, $\forall r \in [0, \rho(0))$ and any m, $\alpha(r, m) = a^R(0)$.*

Proof. Pick any $\hat{r} < \rho(0)$ and any m. We claim that $\Phi(0, \hat{r}, m) = 1$, which suffices to prove the Lemma. Let $a_0 \equiv \alpha(\rho(0), \mu(0))$. Since $\xi_l(\hat{r}) = a^R(0)$, and $\xi_h(\hat{r}) = a_0$, it suffices to show that $\forall a \in \left[a^R(0), a_0\right]$, $\forall t > 0$,

$$\begin{array}{rcl} U(a, t) - kC(\hat{r}, t) & \geq & U(\alpha(\rho(t), \mu(t)), t) - kC(\rho(t), t) \\ & \Downarrow & \\ U(a, 0) - kC(\hat{r}, 0) & > & U(a_0, 0) - kC(\rho(0), 0) \end{array}$$

Equilibrium requires

$$U(\alpha(\rho(t), \mu(t)), t) - kC(\rho(t), t) \geq U(a_0, t) - kC(\rho(0), t)$$

and hence it suffices to show that

$$\begin{aligned} & U(a,0) - kC(\hat{r},0) - [U(a_0,0) - kC(\rho(0),0)] \\ > \; & U(a,t) - kC(\hat{r},t) - [U(a_0,t) - kC(\rho(0),t)] \end{aligned}$$

This inequality can be rewritten as

$$\int_a^{a_0} \int_0^t U_{12}(y,z)\, dzdy > k \int_{\hat{r}}^1 \int_0^t C_{12}(y,z)\, dzdy$$

which holds because $C_{12} < 0 < U_{12}$. □

Proof of Lemma 4 on Page 15. Part 1 was proved in Lemma A.1, and Part 3 in Lemma A.2. We prove part 2 now. Suppose not, so that there is pooling on $r_p < \rho(1)$. Obviously $r_p < 1$. We'll write $t_h(r_p)$ as just t_h and $t_l(r_p)$ as just t_l, to reduce notation. There are two conceptually different cases: either $t_h < 1$, or $t_h = 1$.

Case 1: $t_h < 1$

Note that by definition, for small enough $\varepsilon > 0$, $\rho(t_h - \varepsilon) = r_p$, whereas for all $\varepsilon > 0$, $\rho(t_h + \varepsilon) > r_p$. Clearly, for any m, $\alpha(r_p, m) \in \left(a^R(t_l), a^R(t_h)\right)$. Lemma A.1 established that there is $\theta > 0$ such that reports in $(r_p, r_p + \theta)$ are not played in equilibrium. So pick any $\hat{r} \in (r_p, r_p + \theta)$. Report monotonicity implies that for any m, $\alpha(r_p + \theta, m) \geq \alpha(\hat{r}, m) \geq \alpha(r_p, m)$. We will prove that the mD1 criterion requires $\alpha(\hat{r}) = a^R(t_h)$ via two Claims.

Let $\rho^+ \equiv \inf_{t>t_h} \rho(t)$, $\mu^+ \equiv \inf_{t>t_h} \mu(t)$, and $\mu_- \equiv \sup_{t<t_h} \mu(t)$. Note that all three are well-defined by report monotonicity and the fact that bunches are connected. Obviously, $\mu(t_h - \varepsilon) = \mu_-$ for small enough $\varepsilon > 0$.

Claim: Type t_h is indifferent between playing (ρ^+, μ^+), $(\rho(t_h), \mu(t_h))$, and (r_p, μ_-).

Proof: We first prove indifference between $(\rho(t_h), \mu(t_h))$ and (r_p, μ_-). Suppose not, by way of contradiction. Equilibrium implies

$$U(\alpha(\rho(t_h), \mu(t_h)), t_h) - kC(\rho(t_h), t_h) > U(\alpha(r_p, \mu_-), t_h) - kC(r_p, t_h)$$

But then by continuity of U and C, for small enough $\varepsilon > 0$, a type $t_h - \varepsilon$ would rather play $(\rho(t_h), \mu(t_h))$ than (r_p, μ_-), which contradicts equilibrium.

Next we prove indifference between (ρ^+, μ^+) and (r_p, μ_-). Define

$$W(\varepsilon) = U\left(\alpha\left(\rho^+, \mu^+\right), t_h - \varepsilon\right) - kC\left(\rho^+, t_h - \varepsilon\right) - \left[U(\alpha(r_p, \mu_-), t_h - \varepsilon) - kC(r_p, t_h - \varepsilon)\right]$$

Note that W is continuous since U and C are. If t_h is not indifferent between (ρ^+, μ^+) and (r_p, μ_-), then $W(0) \neq 0$. Consider first $W(0) > 0$. Then by continuity, $W(\varepsilon) > 0$ for small enough $\varepsilon > 0$, so that a type $t_h - \varepsilon$ would strictly prefer to play (ρ^+, μ^+) rather than (r_p, μ_-), contradicting equilibrium. If $W(0) < 0$, then type t_h strictly prefers to play (ρ^+, μ^+) rather than (r_p, μ_-). But we already showed that t_h is indifferent between playing $(\rho(t_h), \mu(t_h))$ and (r_p, μ_-). So t_h strictly prefers (ρ^+, μ^+) over $(\rho(t_h), \mu(t_h))$, contradicting equilibrium. ‖

Claim: For all $r \in (r_p, r_p + \theta)$ and all m, $\alpha(r, m) = a^R(t_h)$.

Proof: Pick any $\hat{r} \in (r_p, r_p + \theta)$. Note that $\xi_l(\hat{r}) = \alpha(r_p, \mu_-)$ and $\xi_h(\hat{r}) = \alpha(\rho^+, \mu^+)$. Therefore, we must show that $\forall a \in [\alpha(r_p, \mu_-), \alpha(\rho^+, \mu^+)]$, $\forall t \neq t_h$,

$$\begin{aligned} U(a,t) - kC(\hat{r},t) \;&\geq\; U(\alpha(\rho(t), \mu(t)), t) - kC(\rho(t), t) && \text{(A-1)} \\ &\Downarrow \\ U(a,t_h) - kC(\hat{r},t_h) \;&>\; U(\alpha(\rho(t_h), \mu(t_h)), t_h) - kC(\rho(t_h), t_h) && \text{(A-2)} \end{aligned}$$

Consider first $t < t_h$. Equilibrium requires

$$U\left(\alpha\left(\rho\left(t\right),\mu\left(t\right)\right),t\right) - kC\left(\rho\left(t\right),t\right) \geq U\left(\alpha\left(r_p,\mu_-\right),t\right) - kC\left(r_p,t\right)$$

and we know that t_h is indifferent between $\left(\rho\left(t_h\right),\mu\left(t_h\right)\right)$ and $\left(r_p,\mu_-\right)$. So it suffices to show that

$$\begin{array}{rcl} U\left(a,t\right) - kC\left(\hat{r},t\right) & \geq & U\left(\alpha\left(r_p,\mu_-\right),t\right) - kC\left(r_p,t\right) \\ & \Downarrow & \\ U\left(a,t_h\right) - kC\left(\hat{r},t_h\right) & > & U\left(\alpha\left(r_p,\mu_-\right),t_h\right) - kC\left(r_p,t_h\right) \end{array}$$

This is true if

$$\begin{array}{rl} & U\left(a,t_h\right) - kC\left(\hat{r},t_h\right) - \left[U\left(\alpha\left(r_p,\mu_-\right),t_h\right) - kC\left(r_p,t_h\right)\right] \\ > & U\left(a,t\right) - kC\left(\hat{r},t\right) - \left[U\left(\alpha\left(r_p,\mu_-\right),t\right) - kC\left(r_p,t\right)\right] \end{array}$$

which can be be rewritten as

$$\int_{\alpha(r_p,\mu_-)}^{a}\int_{t}^{t_h} U_{12}\left(y,z\right)dzdy > k\int_{r_p}^{\hat{r}}\int_{t}^{t_h} C_{12}\left(y,z\right)dzdy$$

which is true because $U_{12} > 0 > C_{12}$.

Now consider the other case, $t > t_h$. Equilibrium requires

$$U\left(\alpha\left(\rho\left(t\right),\mu\left(t\right)\right),t\right) - kC\left(\rho\left(t\right),t\right) \geq U\left(\alpha\left(\rho^+,\mu^+\right),t\right) - kC\left(\rho',t\right)$$

and we know that t_h is indifferent between $\left(\rho\left(t_h\right),\mu\left(t_h\right)\right)$ and $\left(\rho^+,\mu^+\right)$. So to show that (A-2) follows from (A-1), it suffices to show that

$$\begin{array}{rcl} U\left(a,t\right) - kC\left(\hat{r},t\right) & \geq & U\left(\alpha\left(\rho^+,\mu^+\right),t\right) - kC\left(\rho^+,t\right) \\ & \Downarrow & \\ U\left(a,t_h\right) - kC\left(\hat{r},t_h\right) & > & U\left(\alpha\left(\rho^+,\mu^+\right),t_h\right) - kC\left(\rho^+,t_h\right) \end{array}$$

This is true if

$$\begin{array}{rl} & U\left(a,t_h\right) - kC\left(\hat{r},t_h\right) - \left[U\left(\alpha\left(\rho^+,\mu^+\right),t_h\right) - kC\left(\rho^+,t_h\right)\right] \\ > & U\left(a,t\right) - kC\left(\hat{r},t\right) - \left[U\left(\alpha\left(\rho^+,\mu^+\right),t\right) - kC\left(\rho^+,t\right)\right] \end{array}$$

which can be rewritten as

$$\int_{a}^{\alpha\left(\rho^+,\mu^+\right)}\int_{t_h}^{t} U_{12}\left(y,z\right)dzdy > k\int_{\hat{r}}^{\rho^+}\int_{t_h}^{t} C_{12}\left(y,z\right)dzdy$$

which is true because $U_{12} > 0 > C_{12}$. ‖

This complete the proof for $t_h < 1$.

Case 2: $t_h = 1$

If $\rho\left(1\right) > r_p$, then the same arguments as in Case 1 work, except that we now define $\rho^+ \equiv \rho\left(1\right)$, and $\mu^+ \equiv \mu\left(1\right)$. So consider $\rho\left(1\right) = r_p < 1$. Pick any $\hat{r} > r_p$. Since $\xi_h\left(\hat{r}\right) = a^R\left(1\right)$, we must show that $\forall a \in \left[\alpha\left(r_p, \mu\left(1\right)\right), a^R\left(1\right)\right], \forall t < 1$,

$$\begin{array}{rcl} U\left(a,t\right)-kC\left(\hat{r},t\right) & \geq & U\left(\alpha\left(\rho\left(t\right),\mu\left(t\right)\right),t\right)-kC\left(\rho\left(t\right),t\right) \\ & \Downarrow & \\ U\left(a,1\right)-kC\left(\hat{r},1\right) & > & U\left(\alpha\left(r_p,\mu\left(1\right)\right),1\right)-kC\left(r_p,1\right) \end{array}$$

Equilibrium requires that for all t,

$$U\left(\alpha\left(\rho\left(t\right),\mu\left(t\right)\right),t\right)-kC\left(\rho\left(t\right),t\right) \geq U\left(\alpha\left(r_p,\mu\left(1\right)\right),t\right)-kC\left(r_p,t\right)$$

So it suffices to show that for all $t < 1$,

$$\begin{array}{rl} & U\left(a,1\right)-kC\left(\hat{r},1\right)-\left[U\left(\alpha\left(r_p,\mu\left(1\right)\right),1\right)-kC\left(r_p,1\right)\right] \\ > & U\left(a,t\right)-kC\left(\hat{r},t\right)-\left[U\left(\alpha\left(r_p,\mu\left(1\right)\right),t\right)-kC\left(r_p,t\right)\right] \end{array}$$

This can be rewritten as

$$\int_{\alpha(r_p,\mu(1))}^{a}\int_t^1 U_{12}\left(y,z\right)dzdy > k\int_{r_p}^{\hat{r}}\int_t^1 C_{12}\left(y,z\right)dzdy$$

This inequality holds because $C_{12} < 0 < U_{12}$. □

Lemma A.3. *In any monotonic D1 equilibrium, $\rho\left(t\right) > t$ for all $t \in (0,1)$.*

Proof. Denote the cutoff type by $\hat{t} < 1$. First, we prove the Lemma for $t \in \left(0,\hat{t}\right)$. Suppose not. It is straightforward to check that Lemma 1 implies that $\rho\left(t\right) \neq t$ for all $t \in \left(0,\hat{t}\right)$. So there exists $t' \in \left(0,\hat{t}\right)$ such that $\rho\left(t'\right) < t'$. By report monotonicity, $\rho\left(t'-\varepsilon\right) < \rho\left(t'\right)$ for all $\varepsilon > 0$. It follows that for small enough $\varepsilon > 0$, $C\left(\rho\left(t'-\varepsilon\right),t'-\varepsilon\right) > C\left(\rho\left(t'\right),t'-\varepsilon\right)$. On the other hand, since we are on the separating part of the type space, $U\left(a^R\left(t'\right),t'-\varepsilon\right) > U\left(a^R\left(t'-\varepsilon\right),t'-\varepsilon\right)$. Therefore, a type $t'-\varepsilon$ strictly prefers to imitate t', contradicting equilibrium separation.

Next, observe that $1 = \rho\left(t\right) > t$ for all $t \in \left(\hat{t},1\right)$. So it only remains to prove $\rho\left(\hat{t}\right) > \hat{t}$. If $\hat{t}$ is part of the top pool then $\rho\left(\hat{t}\right) = 1 > \hat{t}$, and we are done. If $\hat{t}$ is not part of the pool, then there is separation up to and including $\hat{t}$, and Lemma 1 implies that $\rho\left(\hat{t}\right) \neq \hat{t}$. Report monotonicity and that $\rho\left(t\right) > t$ for all $t \in \left(0,\hat{t}\right)$ then implies that $\rho\left(\hat{t}\right) > \hat{t}$. □

Lemma A.4. *In any mD1 equilibrium with cutoff $\hat{t}$, (i) ρ is continuous at all $t \neq \hat{t}$; and (ii) if $\hat{t} > 0$, then ρ is either right- or left-continuous at $\hat{t}$.*

Proof. (i) We prove the continuity at all $t \neq \hat{t}$ first. Trivially, ρ is continuous above $\hat{t}$. Suppose towards a contradiction that there is a discontinuity at some $t' < \hat{t}$. First assume $\rho\left(t'\right) < \lim_{t \downarrow t'}\rho\left(t\right) \equiv \overline{\rho}$. By the continuity of C and the monotonicity of ρ, as $\varepsilon \searrow 0$,

$$\begin{array}{rcl} C\left(\rho\left(t'+\varepsilon\right),t'+\varepsilon\right)-C\left(\rho\left(t'\right),t'+\varepsilon\right) & \to & C\left(\overline{\rho},t'\right)-C\left(\rho\left(t'\right),t'\right) \\ & > & 0 \end{array}$$

where the inequality follows from $\overline{\rho} > \rho\left(t'\right) \geq t'$. On the other hand, since we are on the separating portion of the type space,

$$U\left(a^R\left(t'+\varepsilon\right),t'+\varepsilon\right)-U\left(a^R\left(t'\right),t'+\varepsilon\right) \to 0$$

Therefore, for small enough $\varepsilon > 0$, $t' + \varepsilon$ prefers to imitate t', contradicting equilibrium separation.

The argument for the other case where $\rho(t') > \lim_{t \uparrow t'} \rho(t)$ is similar, establishing that t' prefers to imitate $t' - \varepsilon$ for small enough $\varepsilon > 0$.

(ii) Suppose not. Since ρ is not right-continuous at $\hat{t}$, then type by report monotonicity, $\hat{t}$ is separating. Since ρ is not left-continuous, report monotonicity implies $\rho(\hat{t}) > \lim_{t \uparrow \hat{t}} \rho(t) \equiv \underline{\rho}$. We will argue that $\hat{t}$ prefers to imitate a type $\hat{t} - \varepsilon$ small enough $\varepsilon > 0$, which contradicts equilibrium separation below $\hat{t}$. Suppose not. Then for all $\varepsilon > 0$,

$$U\left(a^R\left(\hat{t} - \varepsilon\right), \hat{t}\right) - kC\left(\rho\left(\hat{t} - \varepsilon\right), \hat{t}\right) \le U\left(a^R\left(\hat{t}\right), \hat{t}\right) - kC\left(\rho\left(\hat{t}\right), \hat{t}\right)$$

Since $\lim_{\varepsilon \downarrow 0} \rho\left(\hat{t} - \varepsilon\right) = \underline{\rho}$, the LHS is converging to $U\left(a^R\left(\hat{t}\right), \hat{t}\right) - kC\left(\underline{\rho}, \hat{t}\right)$. So the above inequality can hold for all $\varepsilon > 0$ only if $C\left(\underline{\rho}, \hat{t}\right) \ge C\left(\rho\left(\hat{t}\right), \hat{t}\right)$. But by Lemma A.3 and the left-discontinuity hypothesis, $\rho\left(\hat{t}\right) > \underline{\rho} \ge \hat{t}$, whence $C\left(\rho\left(\hat{t}\right), \hat{t}\right) > C\left(\underline{\rho}, \hat{t}\right)$, contradiction. □

Proof of Lemma 5 on Page 16. The second part is obvious, so we'll prove the first. Assume $r_1 < 1$. By the Lemma A.4, either $\rho\left(\hat{t}\right) = r_1$ or $\rho\left(\hat{t}\right) = 1$. So suppose first $\rho\left(\hat{t}\right) = r_1$, in which case $\hat{t}$ is separating. Define for $\varepsilon > 0$,

$$W(\varepsilon) \equiv U\left(a^R\left(\hat{t}\right), \hat{t} + \varepsilon\right) - kC\left(r_1, \hat{t} + \varepsilon\right) - \left[U\left(\alpha(1, m_1), \hat{t} + \varepsilon\right) - kC\left(1, \hat{t} + \varepsilon\right)\right]$$

If the Lemma does not hold, then $W(0) > 0$ (equilibrium prevents $W(0) < 0$). But then by continuity of W, a type $\hat{t} + \varepsilon$ would prefer to imitate $\hat{t}$ rather than pool at the top, contradicting equilibrium. It remains to consider $\rho\left(\hat{t}\right) = 1$, in which case $\hat{t}$ is pooling. Note that then $\mu\left(\hat{t}\right) = m_1$.

Claim: For all m, $U\left(\alpha(1, m_1), \hat{t}\right) \ge U\left(\alpha(1, m), \hat{t}\right)$.

Proof: Suppose not. Then there exist some m such that $U\left(\alpha(1, m_1), \hat{t}\right) < U\left(\alpha(1, m), \hat{t}\right)$. By continuity, it follows that there exists $\varepsilon > 0$ such that for all $t \in \left(\hat{t}, \hat{t} + \varepsilon\right)$, $U(\alpha(1, m_1), t) < U(\alpha(1, m), t)$. But then none of these types t can being playing message m_1, contradicting $m_1 = \lim_{t \downarrow \hat{t}} \mu(t)$. ‖

By the Claim, if the Lemma does not hold, it must be that for some m',

$$U\left(\alpha(1, m_1), \hat{t}\right) - kC\left(1, \hat{t}\right) > U\left(\alpha(r_1, m'), \hat{t}\right) - kC\left(r_1, \hat{t}\right)$$

Claim: For all m, $\alpha(r_1, m) = a^R\left(\hat{t}\right)$.

Proof: Suppose not, for some message m. Lemma 2 implies that $\alpha(r_1, m) \ge \alpha(r, m)$ for all $r \le r_1$. Since all types below $\hat{t}$ are separating and using reports smaller that r_1, it follows that $\alpha(r_1, m) > a^R\left(\hat{t}\right)$. But this can only be optimal for the Receiver if she puts positive probability on some type $t > \hat{t}$ when seeing r_1. We claim that this is ruled out by the mD1 criterion. To show this, it suffices to show that for all $a \in \left[a^R\left(\hat{t}\right), \alpha(1, m_1)\right]$ and $t > \hat{t}$

$$\begin{array}{rcl} U(a, t) - kC(r_1, t) & \ge & U(\alpha(1, \mu(t)), t) - kC(1, t) \\ & \Downarrow & \\ U\left(a, \hat{t}\right) - kC\left(r_1, \hat{t}\right) & \ge & U\left(\alpha(1, m_1), \hat{t}\right) - kC\left(1, \hat{t}\right) \end{array}$$

Since equilibrium requires

$$U(\alpha(1, \mu(t)), t) - kC(1, t) \ge U(\alpha(1, m_1), t) - kC(1, t)$$

it is sufficient if the following inequality holds

$$\begin{aligned} & U\left(a,\hat{t}\right)-kC\left(r_1,\hat{t}\right)-\left[U\left(\alpha\left(1,m_1\right),\hat{t}\right)-kC\left(1,\hat{t}\right)\right] \\ > \quad & U\left(a,t\right)-kC\left(r_1,t\right)-\left[U\left(\alpha\left(1,m_1\right),t\right)-kC\left(1,t\right)\right] \end{aligned}$$

This can be rewritten as

$$\int_a^{\alpha(1,m_1)}\int_{\hat{t}}^t U_{12}\left(y,z\right)dzdy > k\int_{\hat{r}}^1\int_{\hat{t}}^t C_{12}\left(y,z\right)dzdy$$

which holds because $U_{12} > 0 > C_{12}$. ‖
So it must be that

$$U\left(\alpha\left(1,m_1\right),\hat{t}\right)-kC\left(1,\hat{t}\right) > U\left(a^R\left(\hat{t}\right),\hat{t}\right)-kC\left(r_1,\hat{t}\right)$$

But then by continuity, for small enough $\varepsilon > 0$,

$$U\left(\alpha\left(1,m_1\right),\hat{t}-\varepsilon\right)-kC\left(1,\hat{t}-\varepsilon\right) > U\left(a^R\left(\hat{t}\right),\hat{t}\right)-kC\left(r_1,\hat{t}\right)$$

Also, by continuity of U and C, and $\lim_{t\uparrow\hat{t}}\rho\left(t\right)\equiv r_1$, we have that as $\varepsilon\searrow 0$,

$$U\left(a^R\left(\hat{t}-\varepsilon\right),\hat{t}-\varepsilon\right)-kC\left(\rho\left(\hat{t}-\varepsilon\right),\hat{t}-\varepsilon\right)\to U\left(a^R\left(\hat{t}\right),\hat{t}\right)-kC\left(r_1,\hat{t}\right)$$

We conclude that for small enough $\varepsilon > 0$,

$$U\left(\alpha\left(1,m_1\right),\hat{t}-\varepsilon\right)-kC\left(1,\hat{t}-\varepsilon\right) > U\left(a^R\left(\hat{t}-\varepsilon\right),\hat{t}-\varepsilon\right)-kC\left(\rho\left(\hat{t}-\varepsilon\right),\hat{t}-\varepsilon\right)$$

implying that a type $\hat{t}-\varepsilon$ prefers playing $(1,m_1)$ rather than separating, contradicting equilibrium. □

Lemma A.5. *There is a unique solution to the differential equation*

$$\rho'\left(t\right)=g\left(\rho,t\right)\equiv\frac{U_1\left(a^R\left(t\right),t\right)a^R\left(t\right)}{kC_1\left(\rho\left(t\right),t\right)}$$

with initial value $\rho\left(0\right)=\varepsilon$.

Proof. Since we can't rely on standard ODE results, we proceed as follows.

Step 1: Local existence.

Set $t_0\equiv 0$ and $r_0\equiv\varepsilon$. Observe that $C_1\left(r_0,t_0\right) > 0$. Since C and U are $\mathcal{C}^2$, g is $\mathcal{C}^1$ in a small neighborhood around (t_0,r_0), and hence satisfies a Lipschitz condition in this neighborhood. By standard local existence theorems (e.g. Coddington and Levinson (1955, Theorem 2.3)), there is a unique solution, call it $\beta\left(t\right)$, satisfying $\beta\left(t_0\right)=r_0$ and defined on some neighborhood $[t_0,t_0+\delta)$, $\delta > 0$. This solution is $\mathcal{C}^1$.

Step 2: Continuation of the solution.

Note that $C_1\left(\beta\left(t\right),t\right) > 0$ for t close enough to t_0 because $C_1\left(r_0,t_0\right) > 0$, and hence $\beta'\left(t\right) > 0$ for $t\in[0,\delta_1)$ and small enough $\delta_1 > 0$. To prove that a unique extension of β to $[0,1]$ exists, it is sufficient to prove an inductive step that given a solution $\beta\left(t\right)$ on $[0,\delta)$, $\delta > 0$, satisfying $\beta'\left(t\right) > 0$, we can extend it uniquely to $[0,\delta+\theta)$ for some $\theta > 0$ while maintaining that it is continuously differentiable and $\beta' > 0$. To see that this is sufficient, suppose that some $\bar{t} < 1$ is the *sup* over all t such that β can be continued to $[0,t)$. This means that β cannot be continued to $\left[0,\bar{t}+\theta\right)$ for any $\theta > 0$, contradicting the inductive step.

The inductive step.

Assume a solution $\beta \in \mathcal{C}^1$ on $[0, \delta)$ with $\beta' > 0$.

Claim: $r_\delta \equiv \lim_{t \uparrow \delta} \beta(t)$ exists and is finite.

Proof: Since β is continuous and $\beta'(t) > 0$ on $[0, \delta)$, it follows immediately that $r_\delta \equiv \lim_{t \uparrow \delta} \beta(t)$ exists and lies in $(r_0, \infty]$. Suppose by way of contradiction that $r_\delta = \infty$. Since β is continuously differentiable on $[0, \delta)$, and $\lim_{t \uparrow \delta} \beta(t) = \infty$ for finite δ, it must be that $\lim_{t \uparrow \delta} \beta'(t) = \infty$. Observe that $a^R(\delta)$ and $U^S\left(a^R(\delta), \delta, b\right)$ are finite (by continuity of the relevant functions), whereas $\lim_{t \uparrow \delta} C_1(\beta(t), t) = \infty$ since $\lim_{t \uparrow \delta} \beta(t) = \infty$ and $C_{11} > 0$. Therefore, we can find a small enough $\varepsilon > 0$ such that

$$\beta'(\delta - \varepsilon) > \frac{U^S\left(a^R(\delta - \varepsilon), \delta - \varepsilon, b\right) a^R(\delta - \varepsilon)}{kC_1(\beta(\delta - \varepsilon), \delta - \varepsilon)}$$

which means that β is not a solution to the differential equation for $t = \delta - \varepsilon$, contradiction. $\|$

Claim: g is continuous in a neighborhood of (δ, r_δ).

Proof: It is sufficient to establish that $C_1(r_\delta, \delta) \neq 0$, since then continuity of g around (δ, r_δ) follows from the continuity of C_1 in a neighborhood around (δ, r_δ), and continuity of U^S and a^R. Suppose $C_1(r_\delta, \delta) = 0$, by way of contradiction. Note that the RHS numerator of the differential equation is continuous in t and strictly positive at $t = \delta$. But then the continuity of β' on $[0, \delta)$ implies that $\lim_{t \uparrow \delta} \beta'(t) = \infty$, which contradicts what we established in the previous Claim. $\|$

Given these two Claims, the argument used in the proof of the Continuation Theorem in Coddington and Levinson (1955, p. 14) allows us to conclude that there is a unique extension, which is $\mathcal{C}^1$, of β to $[0, \delta + \theta)$ for some $\theta > 0$.[37] Note that since the continuation maintains continuity of β', it must be that $\beta' > 0$ on $[0, \delta + \theta)$. $\square$

Proof of Lemma 6 on Page 17. First tackle existence.[38] Consider a perturbed problem defined by a parameter $\varepsilon > 0$:

$$\rho'(t) = \frac{U_1\left(a^R(t), t\right) a^R(t)}{kC_1(\rho(t), t)}, \qquad \rho(0) = \varepsilon$$

[37] The details are as follows. Given the continuity of g around (δ, r_δ), the Picard Local Existence Theorem implies a unique solution, call it $\zeta(t)$, satisfying $\zeta(\delta) = r_\delta$ and defined on some interval $(\delta - \theta, \delta + \theta)$, $\theta > 0$. So define $\hat{\beta}$ by

$$\hat{\beta}(t) = \begin{cases} \beta(t) \text{ if } t \in [0, \delta) \\ r_\delta \text{ if } t = \delta \\ \zeta(t) \text{ if } t \in [\delta, \delta + \beta) \end{cases}$$

To show that $\hat{\beta}$ is a continuously differentiable extension of β to the right of δ, the only point to check is the existence and continuity of the derivative of $\hat{\beta}$ at δ. We claim that for any $t \in [0, \delta + \beta)$

$$\hat{\beta}(t) = r_0 + \int_0^t g\left(z, \hat{\beta}(z)\right) dz$$

This is obvious for $t \in [0, \delta)$. For $t = \delta$, it follows immediately from the definition of r_δ. For $t > \delta$, it follows from the fact that $\zeta(\delta) = r_\delta$ and that $\zeta'(t) = g(t, \zeta(t))$ for $t \geq \delta$. The continuity of $\hat{\beta}$ (which is obvious by construction) implies that of $g\left(z, \hat{\beta}(z)\right)$, and by differentiation of the above integral equation for $\hat{\beta}$, we obtain that $\hat{\beta}'(t) = g\left(t, \hat{\beta}(t)\right)$, whereby $\hat{\beta}'(t)$ exists and is continuous.

[38] As noted in the text, standard results don't apply because we don't have a Lipschitz condition on $(t, \rho) \in [0, 1] \times [0, \infty)$.

where the relevant domain is $t \in [0,1]$ and $\rho \in [\varepsilon, \infty)$. Lemma A.5 proves existence of a unique solution to this initial value problem. Denote this solution as $\sigma_\varepsilon(t)$, and recall that it is continuously differentiable and satisfies $\sigma'_\varepsilon > 0$. Since $\sigma_\varepsilon(0) = \varepsilon > 0$, and $\sigma'_\varepsilon(t) \to \infty$ as $\sigma(t) \to t$, the continuity of $\sigma(t)$ implies that there some $\delta_\varepsilon > 0$ s.t $\forall t \in [0,1]$, $\sigma_\varepsilon(t) \geq t + \delta_\varepsilon$. Observe that the DE is Lipschitz on the restricted domain $\{t, \rho : t \in [0,1], \rho \geq t + \underline{\varepsilon}\}$ for arbitrary $\underline{\varepsilon} > 0$. Therefore, standard arguments imply that $\sigma_\varepsilon(t)$ is continuous in ε for all $\varepsilon > 0$.[39] Consequently, $\sigma_0(t) \equiv \lim_{\varepsilon \downarrow 0} \sigma_\varepsilon(t)$ is well-defined. Since each $\sigma_\varepsilon(t)$ is continuously differentiable, it is immediate that for all $t \in [0,1]$, $\sigma'_0(t)$ is well-defined, with

$$\begin{aligned}
\sigma'_0(t) &= \lim_{\varepsilon \downarrow 0} \sigma'_\varepsilon(t) \\
&= \frac{U_1\left(a^R(t), t\right) a^R(t)}{kC_1\left(\lim_{\varepsilon \downarrow 0} \sigma_\varepsilon(t), t\right)} \\
&= \frac{U_1\left(a^R(t), t\right) a^R(t)}{kC_1\left(\sigma_0(t), t\right)}
\end{aligned}$$

Since $\sigma_0(0) \equiv \lim_{\varepsilon \downarrow 0} \sigma_\varepsilon(0) = 0$, we conclude that σ_0 is a solution to (DE).

To prove uniqueness, suppose by way of contradiction that $\eta \neq \sigma_0$ is a solution to (DE). Note that η must be continuously differentiable. On account of $\eta'(0) = \sigma'_0(0) = \infty$, it must be that there is some $\hat{t} > 0$ such that $\min\{\eta(t), \sigma_0(t)\} > t$ for all $t \in (0, \hat{t})$. Since we can prove that for all $t \in (0, \hat{t})$, there is a unique solution to the DE with initial condition $(t, \eta(t))$, it follows that for all $t \in (0, \hat{t})$, $\eta(t) \neq \sigma_0(t)$. By continuity of both functions, one of them must be strictly above the other in this region, so assume without loss of generality that for all $t \in (0, \hat{t})$, $\eta(t) > \sigma_0(t) > t$. By $C_{11} > 0$, the DE implies that $\eta'(t) < \sigma'_0(t)$ for all $t \in (0, \hat{t})$. But by continuity, this implies $\sigma_0(t) > \eta(0)$ for small t. Contradiction. □

Proof of Theorem 2 on Page 18. The proof is constructive.

Step 0: Preliminaries

Start by defining the function

$$\phi(t) \equiv U\left(a^S(t), t\right) - kC(1, t) - \left[U\left(a^R(t), t\right) - kC(\rho^*(t), t)\right]$$

$\phi(t)$ is the gain for type t from sending the highest report and receiving his ideal action over separating himself (thus inducing $a^R(t)$) with report $\rho^*(t)$. Note that in equilibrium, the gain from pooling over separating can be no more than $\phi(t)$, and will generally be strictly less. Clearly ϕ is continuous, and $\phi(\bar{t}) > 0$. There are two conceptually distinct cases: one where $\phi(t) = 0$ for some $t \leq \bar{t}$, and the other where $\phi(t) > 0$ for all $t \leq \bar{t}$. Define

$$\underline{t} \equiv \begin{cases} 0 & \text{if } \phi(t) > 0 \text{ for all } t \leq \bar{t} \\ \sup_{t \in [0,\bar{t}]}\{t : \phi(t) = 0\} & \text{otherwise} \end{cases}$$

Note that a necessary condition for $\underline{t} = 0$ is that $\phi(0) \geq 0$. In everything that follows, we are mainly concerned with $t \in [\underline{t}, \bar{t}]$. So statements such as "for all t" are to be read as "for all $t \in [\underline{t}, \bar{t}]$" and so forth unless explicitly specified otherwise. Note that for all $t \in (\underline{t}, \bar{t}]$, $\phi(t) > 0$.

Step 1: Constructing the necessary sequences.

[39] See for example Coddington and Levinson (1955, Theorem 7.1, p. 22) or Birkhoff and Rota (1989, Theorem 2, p. 175).

Initialize $p_0^L(t) = p_0^R(t) = t$, and $a_0^L(t) = a_0^R(t) = a^R(t)$. Define

$$\Delta(a,t) \equiv U(a,t) - kC(1,t) - \left[U\left(a^R(t), t\right) - kC(\rho^*(t), t)\right]$$

Clearly, Δ is continuous in both arguments, and strictly concave in a with a maximum at $a^S(t)$. Since $\Delta\left(a^R(t), t\right) \leq 0 \leq \Delta\left(a^S(t), t\right)$ for all $t \in \left[\underline{t}, \overline{t}\right]$, it follows for any relevant t, in the domain $a \in \left[a^R(t), a^S(t)\right]$ there exists a unique solution to $\Delta(a,t) = 0$. Call this $a_1^L(t)$. Similarly, on the domain $a \in \left[a^S(t), \infty\right)$, there exists a unique solution to $\Delta(a,t) = 0$. Call this $a_1^R(t)$. Note that by continuity of Δ, a_1^L and a_1^R are continuous, $a_1^L(\overline{t}) = a_0^R(\overline{t})$, and $a_1^R(\underline{t}) = a_1^L(\underline{t}) = a^S(\underline{t})$ if $\underline{t} > 0$. Since the function $\overline{a}(t_1, t_2)$ is strictly increasing in both arguments for $t_1, t_2 \in [0,1]$ and constant outside $[0,1]$, given t there is either no or a unique t' that solves $\overline{a}(t, t') = a_1^q(t)$ for $q \in \{L, R\}$. If there is a solution, call it $p_1^q(t)$, otherwise set $p_1^q(t) = 1$ (for each $q \in \{L, R\}$). It follows that that p_1^L and p_1^R are continuous functions, $p_1^L(t) \geq p_0^L(t)$ with equality if and only if $t = \overline{t}$, and $p_1^R(t) > p_0^R(t)$. Note that $p_1^R(t) \geq p_1^L(t)$, and $p_1^L(\underline{t}) = p_1^R(\underline{t})$ if $\underline{t} > 0$.

For $j \geq 2$ and $q \in \{L, R\}$, recursively define $p_j^q(t)$ as the solution to

$$U\left(\overline{a}\left(p_{j-1}^q(t), p_j^q(t)\right), p_{j-1}^q(t)\right) - U\left(\overline{a}\left(p_{j-2}^q(t), p_{j-1}^q(t)\right), p_{j-1}^q(t)\right) = 0$$

if a solution exists that is strictly greater than $p_{j-1}^q(t)$, and otherwise set $p_j^q(t) = 1$. By the monotonicity (constancy) of $\overline{a}$ in (outside) the type space, and concavity of U in the first argument, $p_j^q(t)$ is well-defined and unique. Define $a_j^q(t) \equiv \overline{a}\left(p_{j-1}^q(t), p_j^q(t)\right)$. Note that for all $j \geq 2$, $p_j^q(t) > p_{j-1}^q(t)$ and $a_j^q(t) > a_{j-1}^q(t)$ if and only if $p_{j-1}^q(t) < 1$. For all j and $q \in \{L, R\}$, $p_j^q(t)$ is continuous, $p_j^R(t) \geq p_j^L(t)$ for all t, $p_j^L(\underline{t}) = p_j^R(\underline{t})$ if $\underline{t} > 0$, and $p_{j+1}^L(\overline{t}) = p_j^R(\overline{t})$ (these follow easily by induction, given that we noted all these properties for $j = 1$).

Step 2: The critical segment M

I now argue that exists $M \geq 1$ such that $p_{M-1}^R(\overline{t}) < 1 = p_M^R(\overline{t})$. (Obviously, if it exists, it is unique.) To prove this, first note that by definition, $p_0^R(\overline{t}) = \overline{t} < 1$. Let $\overline{K} = \inf\{K : p_K^R(\overline{t}) = 1\}$.[40] It is sufficient to show that $\exists \varepsilon > 0$ such that for any $j \in \{0, \ldots, \overline{K} - 1\}$, $\left|a_{j+1}^R(\overline{t}) - a_j^R(\overline{t})\right| \geq \varepsilon$. For any $j \in \{0, \ldots, \overline{K} - 1\}$, type $p_j^R(\overline{t})$ is indifferent between $a_j^R(\overline{t})$ and $a_{j+1}^R(\overline{t})$, by construction. Since $a_j^R(\overline{t}) < a_{j+1}^R(\overline{t})$, it must be that $a_j^R(\overline{t}) < a^S\left(p_j^R(\overline{t})\right) < a_{j+1}^R(\overline{t})$. On the other hand, by their definitions, we also have $a_j^R(\overline{t}) \leq a^R\left(p_j^R(\overline{t})\right) \leq a_{j+1}^R(\overline{t})$. Since there is a uniform lower bound $\lambda > 0$ on $\left|a^S(t) - a^R(t)\right|$, it follows that $\left|a_{j+1}^R(\overline{t}) - a_j^R(\overline{t})\right| \geq \lambda > 0$ for all $j \in \{0, \ldots, \overline{K} - 1\}$.

Step 3: Existence when $\underline{t} > 0$.

Consider the functions p_M^L and p_M^R. These are continuous, and $p_M^L(\overline{t}) = p_{M-1}^R(\overline{t}) < 1 = p_M^R(\overline{t})$. Moreover, $p_M^L(\underline{t}) = p_M^R(\underline{t})$; hence either $p_M^R(\underline{t}) < 1$ or $p_M^L(\underline{t}) = 1$. It follows that there is some type $\hat{t} \in [\underline{t}, \overline{t}]$ such that either (i) $p_M^L(\hat{t}) = 1$ and $p_M^L(t) < 1$ for all $t > \hat{t}$; or (ii) $p_M^R(\hat{t}) = 1$ and $p_M^R(t) < 1$ for all $t < \hat{t}$. By construction, there is an mD1 equilibrium where all types $t \in \left[0, \hat{t}\right)$ play $\rho^*(t)$, and all types $t \in \left[\hat{t}, 1\right]$ play $\rho(t) = 1$, and further segment themselves using the cheap-talk messages into the partition $\left\{\hat{t}, p_1^q\left(\hat{t}\right), \ldots, p_M^q\left(\hat{t}\right)\right\}$.

Step 4: Existence when $\underline{t} = 0$.

By the continuity of p_M^L and p_M^R, the logic in Step 3 can fail when $\underline{t} = 0$ only if $p_M^L(0) < 1 = p_M^R(0)$. So suppose this is the case. Note that this requires $p_1^L(0) < p_1^R(0)$. For any $t \in \left[p_1^L(0), p_1^R(0)\right]$,

$$U(\overline{a}(0,t), 0) - kC(1,0) - \left[U\left(a^R(0), 0\right) - kC(0,0)\right] \geq 0$$

[40] Recall that the infimum of an empty set is $+\infty$.

with strict inequality for interior t. In words, when $t \in \left[p_1^L(0), p_1^R(0)\right]$, type 0 weakly prefers (indifference at the endpoints and strict preference for interior t) inducing $\overline{a}(0,t)$ with report 1 over inducing $a^R(0)$ with report 0. This follows from the construction of p_1^L and p_1^R. Given any $t \in [0,1]$, define $\tau_0(t) = 0$, $\tau_1(t) = t$, and recursively, for $j \geq 2$, $\tau_j(t)$ as the solution to

$$U\left(\overline{a}\left(\tau_{j-1}(t), \tau_j(t)\right), \tau_{j-1}(t)\right) - U\left(\overline{a}\left(\tau_{j-2}(t), \tau_{j-1}(t)\right), \tau_{j-1}(t)\right) = 0$$

if a solution exists that is strictly greater than $\tau_{j-1}(t)$, and otherwise set $\tau_j(t) = 1$. By the monotonicity (constancy) of $\overline{a}$ in (outside) the type space, and concavity of U in the first argument, $\tau_j(t)$ is well-defined and unique for all j. It is straightforward that for all $j \geq 0$, $\tau_j(t)$ is continuous in t. Since

$$\tau_M\left(p_1^L(0)\right) = p_M^L(0) < 1 = p_M^R(0) = \tau_M\left(p_1^R(0)\right)$$

it follows that

$$\tilde{t} = \min_{t \in \left[p_1^L(0), p_1^R(0)\right]} \{t : \tau_M(t) = 1\}$$

is well-defined and lies in $\left(p_1^L(0), p_1^R(0)\right]$. By construction, there is an mD1 equilibrium where all types send the costly report of 1, and segment themselves using cheap talk messages into the partition $\left\{0 = \tau_0\left(\tilde{t}\right), \tau_1\left(\tilde{t}\right), \ldots, \tau_M\left(\tilde{t}\right)\right\}$. □

Proof of Theorem 3 on Page 22. The first two parts were proved in the text. It remains to show that under Condition M, a CS equilibrium with first segment $[0, t^*]$ is robust. For there to be a CS equilibrium with first segment $[0, t^*]$, it must be that $t^* \leq 1$, so assume that holds in what follows.

Some preliminaries are needed. First, extend the notation in the proof of Theorem 2 in the obvious way to be explicit about the dependence of various objects on k. This defines $\phi(t,k)$, $\overline{t}(k)$, $\underline{t}(k)$, and for $j \in \{0, 1, \ldots\}$, the sequences $p_j^L(t,k)$, $p_j^R(t,k)$. Note that the functions τ_j do *not* depend on k. Finally, note that the critical segment M defined in Step 2 of in the proof of Theorem 2 does depend generally on k and $\overline{t}$, but picking sufficiently small k, $\overline{t}$ can be made arbitrarily close to 0, and hence by continuity of U and C, once k is small enough, M is independent of k and $\overline{t}$; it is equal to the the smallest integer $n \geq 1$ such that $\tau_{n+1}(0) = 1$ and $\tau_n(0) < 1$ (recall that $\tau_0(t) = 0$ and $\tau_1(t) = t$).

Now observe that for sufficiently small k, $\underline{t}(k) = 0$ because $\phi(t,k) > 0$ for all t, for all k small enough. Hence once k is small enough, $p_j^q(0,k)$ is well-defined for all $j \geq 0$, $q \in \{L, R\}$. From the definition of $p_1^R(0,k)$,

$$U(\overline{a}(0, p_1^R(0,k), 0) - U(a^R(0), 0) = k[C(1,0) - C(0,0)]$$

and $\overline{a}(0, p_1^R(0,k) \geq a^S(0)$. It is straightforward that for any $k > 0$, $p_1^R(0,k) < t^*$, and as $k \to 0$, $p_1^R(0,k) \uparrow t^*$. Next, the construction of p_j^R and τ_j implies that for all j, $p_j^R(0,k) = \tau_j(p_1^R(0,k))$, and in particular, $p_M^R(0,k) = \tau_M(p_1^R(0,k))$.

By the definition of M and t^*, $\tau_M(t^*) = 1$, and given Condition M, $\tau_M(t) < 1$ for all $t < t^*$ (this is because $\tau_M(0) < 1$ and if there exists some $\tilde{t} < t^*$ such that $\tau_M(\tilde{t}) = 1$, by continuity of τ_M there is more than one CS equilibrium with M segments, a contradiction under Condition M). Combined with the previous arguments, this implies that for small k, $p_M^R(0,k) < 1$. The argument in Step 3 of the proof of Theorem 2 then applies to yield an mD1 equilibrium with strictly positive cutoff type, $\hat{t}(k)$, for all k sufficiently small.

The proof is completed by picking any sequence of mD1 equilibria with strictly positive cutoffs as $k \to 0$, and noting that since the cutoffs are converging to 0 and the sequence of the boundaries of the first bunch is converging to t^*, the sequence of mD1 outcomes must be converging to the CS outcome with first segment $[0, t^*]$. □

Proof of Proposition 2 on Page 18. Define $\overline{k}$ as the cost that would make $\phi(0) = 0$ (where ϕ is defined in the proof of Theorem 2), i.e. at $\overline{k}$, type 0 is indifferent between inducing $a^R(0)$ with costly report of 0, and inducing $a^S(0)$ by sending costly report of 1. Explicitly,

$$\overline{k} \equiv \frac{U\left(a^S(0), 0\right) - U\left(a^R(0), 0\right)}{C(1,0) - C(0,0)}$$

Since $k > \overline{k}$ implies $\phi(0) < 0$, we have $\underline{t} > 0$ in the proof of Theorem 2, and hence $p_1^L(\underline{t}) = p_1^R(\underline{t})$, which is sufficient for existence of equilibrium with cutoff strictly above 0. □

Proof of Lemma 8 on Page 24. **Existence:** If $t^* \geq 1$, then $t_1^0 \leq t^*$ for the babbling equilibrium, and we are done. So assume henceforth $t^* < 1$. By its definition, note also that $t^* > 0$.

For $t \in [0, t^*]$, let $p_0(t) = 0$, $p_1(t) = t$, and $a_1(t) = \overline{a}(0, t)$. For $j \geq 2$, recursively define $p_j(t)$ as the solution to

$$U(\overline{a}(p_{j-1}(t), p_j(t)), p_{j-1}(t)) - U(\overline{a}(p_{j-2}(t), p_{j-1}(t)), p_{j-1}(t)) = 0$$

if a solution exists that is strictly greater than $p_{j-1}(t)$, and otherwise set $p_j(t) = 1$. By the monotonicity (constancy) of $\overline{a}$ in (outside) the type space, and concavity of U in the first argument, $p_j(t)$ is well-defined and unique for all j. Define for all $j \geq 1$, $a_j(t) \equiv \overline{a}(p_{j-1}(t), p_j(t))$. One can show by induction that p_j is continuous for all $j \geq 1$. Since $p_2(0) = t^*$, induction also yields that $p_j(0) = p_{j-1}(t^*)$ for all $j \geq 1$.

It can be shown that there exists an $M \geq 1$ such that $p_{M-1}(t^*) < 1 = p_M(t^*)$. (Obviously, given that it exists, it is unique.) The proof is analogous to Step 2 of Theorem 2. This implies that $p_M(0) = p_{M-1}(t^*) < 1 = p_M(t^*)$. Since p_M is continuous, it follows that

$$\tilde{t} = \min_{t \in [0, t^*]} \{t : p_M(t) = 1\}$$

is well-defined and lies in $(0, t^*]$. By construction, the partition

$$\left\{0 = p_0\left(\tilde{t}\right), \tilde{t} = p_1\left(\tilde{t}\right), \ldots, p_M\left(\tilde{t}\right) = 1\right\}$$

is a CS equilibrium with first segment boundary $\tilde{t} \leq t^*$.

Uniqueness under Condition M: Now assume Condition M. We must show that there is a unique CS outcome with first segment $t_1^0 \leq t^*$. First note that under Condition M, if $t^* \geq 1$, then babbling is the unique CS outcome; hence, assume that $t^* < 1$. The key observation is that under Condition M, for all $i \geq 0$, the function $p_i(t)$ is non-decreasing everywhere, and strictly increasing if $p_i(t) < 1$. This follows from the construction of p_i and the definition of Condition M. Therefore, since $p_n(0) < 1$ and $p_n(t^*) < 1$ for all $n < M$, there is no n-segment CS outcome with first segment boundary $t_1 \leq t^*$ for any $n < M$. Crawford and Sobel (1982, Lemma 2) proved that under Condition M, CS outcomes with more segments have shorter first segments. Accordingly, it suffices to show that there is no $(M+1)$-segment CS outcome. But this follows from the facts that $p_{M+1}(0) = p_M(t^*) = 1$ and p_{M+1} is non-decreasing; hence there is no $t' \in (0, \tilde{t})$ s.t.

$$t' = \min_{t \in (0, t^*]} \{t : p_{M+1}(t) = 1\}$$

and thus no $(M+1)$-segment CS outcome. □

Proof of Proposition 6 on Page 28. It has already been proved in Theorem 3 that babbling can be supported for small enough k in an mD1, and hence D1, equilibrium if $U(\overline{a}(0,1),0) > U(a^R(0),0)$, so assume henceforth that $U(\overline{a}(0,1),0) \leq U\left(a^R(0),0\right)$, which in the U-Q model is equivalent to the bias parameter b satisfying $b \leq \frac{1}{4}$. Also, define $a_u \equiv \overline{a}(0,1) = \frac{1}{2}$. Define r_u by $U_1(a_u, r_u) = 0$, so that $r_u = a_u - b$. I will argue that a D1 equilibrium exists where all types pool on the costly report of r_u and all send the same cheap talk message.

Claim: For all $t \leq r_u$, if $(a,r) \succeq_t (a_u, r_u)$ then $(a,r) \succeq_0 (a_u, r_u)$.

Proof: We must show that if $t \leq r_u$, then

$$\begin{aligned} -(a-t-b)^2 - k(r-t)^2 &\geq -(a_u-t-b)^2 - k(r_u-t)^2 \\ &\Downarrow \\ -(a-b)^2 - kr^2 &\geq -(a_u-b)^2 - kr_u^2 \end{aligned} \tag{A-3}$$

It suffices to show that whenever (A-3) holds,

$$(a_u-b)^2 - (a-b)^2 - k\left(r^2 - r_u^2\right) \geq (a_u-t-b)^2 - (a-t-b)^2 - k\left((r-t)^2 - (r_u-t)^2\right)$$

which after some algebra simplifies to

$$a_u - a \geq k(r - r_u) \tag{A-4}$$

After some algebra, we can also rewrite inequality (A-3) as

$$[a - b + r_u - 2t](a_u - a) \geq [r + r_u - 2t]\,k(r - r_u) \tag{A-5}$$

So it suffices to show that (A-5)$\Rightarrow$(A-4). We'll break it into three cases.

Case 1: $r \geq r_u$. Then (A-5) can only hold if $a_u \geq a$ (given that $t \leq r_u = a_u - b$), whence it suffices to show that

$$r_u + a - 2t - b \leq r + r_u - 2t$$

This holds because $r \geq r_u = a_u - b \geq a - b$.

Case 2: $r < r_u$ and $a \leq a_u$. Then (A-4) trivially holds since the LHS is weakly positive, whereas the RHS is strictly negative.

Case 3: $r < r_u$ and $a > a_u$. Then it is easy to see that $a - b + r_u - 2t > 0$ (since $a - b > a_u - b = r_u \geq t$) and therefore (A-5) can only hold if $r + r_u - 2t > 0$. But then, (A-5) is equivalent to $k\frac{r_u - r}{a - a_u} \geq \frac{a-b+r_u-2t}{r+r_u-2t} \geq 1$ (since $a - b > a_u - b = r_u > r$) and (A-4) holds. $\|$

Claim: For all $t \geq r_u$, if $(a,r) \succeq_t (a_u, r_u)$ then $(a,r) \succeq_1 (a_u, r_u)$.

Proof: Similar to above. $\|$

The two claims together show that in (a,r) space, the indifference curve for any type $t \in [0, r_u]$ is weakly "to the left" of that of type 0, and the indifference curve for any type $t \in [r_u, 1]$ is weakly "to the right" of that of type 1. Now I show that the indifference curve for type 1 is *strictly* "to the right" of the indifference curve for type 0.

Claim: For any $r \neq r_u$, if $(a,r) \succeq_1 (a_u, r_u)$ then $(a,r) \prec_0 (a_u, r_u)$.

Proof: With some algebra, $(a,r) \succeq_1 (a_u, r_u)$ is equivalent to

$$(2 - r_u - a + b)(a_u - a) \leq k(r - r_u)(2 - r_u - r) \tag{A-6}$$

and $(a,r) \prec_0 (a_u, r_u)$ is equivalent to

$$(r_u + a - b)(a_u - a) < k(r + r_u)(r - r_u) \tag{A-7}$$

Since $(a,r)\,(a_u, r_u)$ is true if $r > r_u$ and $a > a_u$, and conversely $(a,r) \succeq_1 (a_u, r_u)$ is not true if $a < a_u$ and $r < r_u$, we only need to show that (A-6)⇒(A-7) in two regions: $\{r > r_u, a \leq a_u\}$ and $\{r < r_u, a \geq a_u\}$.

Case 1: $r > r_u, a \leq a_u$. Then (A-6) implies

$$\begin{aligned} \frac{1}{k} &\leq \left(\frac{r - r_u}{a_u - a}\right)\left(\frac{2 - r_u - r}{2 - r_u - a + b}\right) \\ &< \frac{r - r_u}{a_u - a} \\ &\leq \left(\frac{r - r_u}{a_u - a}\right)\left(\frac{r + r_u}{r_u + a - b}\right) \end{aligned}$$

where the second inequality follows from $r > r_u = a_u - b \geq a - b$, and the third follows from $r_u = a_u - b \geq a - b$. So (A-7) holds.

Case 2: $r < r_u, a \geq a_u$. Then (A-6) implies

$$\begin{aligned} \frac{1}{k} &\geq \left(\frac{r_u - r}{a - a_u}\right)\left(\frac{2 - r_u - r}{2 - r_u - a + b}\right) \\ &> \frac{r_u - r}{a - a_u} \\ &\geq \left(\frac{r_u - r}{a - a_u}\right)\left(\frac{r + r_u}{r_u + a - b}\right) \end{aligned}$$

where the second inequality follows from $r < r_u = a_u - b \leq a - b$, and the third follows from $r_u = a_u - b \leq a - b$. So (A-7) holds. ‖

Note that the contrapositive to this Claim establishes that for all $r \neq r_u$, if $(a, r) \succeq_0 (a_u, r_u)$ then $(a, r) \prec_1 (a_u, r_u)$. Now observe that when k is small enough (and up to now we haven't used the smallness of k yet), for all $r \neq r_u$, there exists some $a_0(r)$ and $a_1(r)$ such that $(a_t(r), r)$ is strictly preferred to (a_u, r_u) by type $t \in \{0, 1\}$.[41] By continuity of preferences and the previous claims, it follows that for all $r \neq r_u$, there exists some $a^*(r)$ such that $(a^*(r), r) \prec_t (a_u, r_u)$ for all $t \in [0, 1]$. Set $a^*(r_u) = a_u$. Pick some cheap talk message m_u. I claim that $\rho(t) = r_u$ and $\mu(t) = m_u$ for all t and $\alpha(r, m) = a^*(r)$ is a D1 equilibrium. By construction, the Sender is playing optimally. The D1 criterion cannot rule out putting probability on type 0 or type 1 since by the arguments above, there is no type who strictly prefers deviating for every response than either type 0 or type 1 weakly prefers deviating for; hence the Receiver can hold any beliefs that only puts probability on these two types.[42] Since each $a^*(r)$ can be rationalized by some mixture of beliefs over these two types, the strategies constitute a D1 equilibrium.

□

[41] If k is too large, then for instance type 1 will prefer $(a, 1)$ to (a_u, r_u) for any $a \in [0, 1]$, which means that babbling on r_u cannot be an equilibrium.

[42] Though the Claims didn't assert this, one can further check that the Receiver cannot put positive probability on any interior type under D1.

Appendix B: Unbounded Type Space

The body of the paper assumed that the type space is bounded, normalized to $T \equiv [0,1]$. In this Appendix, I discuss briefly what happens if instead the type space is unbounded, so that $T \equiv (-\infty, \infty)$.[43] Assume that the report space is also $\mathcal{R} \equiv (-\infty, \infty)$. CS only characterize their pure cheap talk model for a bounded type space. However, with an added technical condition, it is easy to verify that their main results go through when communication is through cheap talk alone.

Assumption B.1. *There exists $\lambda > 0$ such that for all $t \in T$, $a^S(t) - a^R(t) > \lambda$.*

Note that this is automatically satisfied with a bounded type space. Given this, we can state

Lemma B.1. ***(CS Lemma 1)*** *Assume $k = 0$. Then there exists $\nu > 0$ such that if a_1 and a_2 are two actions played in equilibrium, $\mid a_1 - a_2 \mid > \nu$.*

The proof is exactly the same as Crawford and Sobel (1982, Lemma 1). Whereas this implies a finite number of actions played in equilibrium with a bounded type space (as in CS), it only implies a countable number of actions played in the unbounded case. Nonetheless, it does imply no full separation.[44]

In contrast, for any $k > 0$, a fully separating equilibrium *does* exist. In fact, it can be constructed so that all the reports are used in equilibrium, so that the mD1 criterion is automatically satisfied.

Proposition B.1. *There is a fully-revealing mD1 equilibrium for any $k > 0$.*

Proof Sketch. The same differential equation as in (DE) applies, but one does not need to impose the initial value condition. So picking any $\varepsilon > 0$ and setting $\rho(0) = \varepsilon$, the argument of Lemma A.5 can be extended to show that there is a solution to the differential equation on the entire domain $r \in (-\infty, \infty)$. It is easy to check that this solution does constitute an equilibrium. Since all reports are used in equilibrium, the mD1 criterion is satisfied.[45] □

Remark B.1. This Proposition is closely related to Ottaviani and Squintani (2004). In a sense, it subsumes Proposition 1 in their paper. In their model, reports are not directly costly, but with some probability β the Receiver's naively plays $a = r$, and with probability $1 - \beta$, she plays the optimal response $a = \alpha(r)$. Accordingly, the Sender's utility from a report r given the Receiver's strategy is

$$(1 - \beta)U(\alpha(r), t) + \beta U(r, t)$$

Compare this with the Sender's utility in the current model (ignoring cheap talk), which is

$$U(\alpha(r), t) - kC(r, t)$$

Clearly, if we set $k \equiv \frac{\beta}{1-\beta}$ and $C(r,t) \equiv -U(r,t)$, then we have the specification of Ottaviani and Squintani (2004). This choice of C does satisfy the assumptions of the model in this paper (where we now no longer have $C_1(t,t) = 0$, but instead $C_1(a^S(t), t) = 0$, which is ok by the discussion in

[43] Of course, one might also want to consider bounded below (above) and unbounded above (below). The analysis is similar.

[44] Moscarini (2004) and Ottaviani and Squintani (2004) also study pure cheap talk models with an unbounded type space and note the no full-separation result.

[45] Details of the proof are available from the author.

Section 7.2.2), and the above Proposition applies to give existence of a fully separating equilibrium, which is what is shown in Proposition 1 of Ottaviani and Squintani (2004). This is not surprising, since the possibility of naivety of the Receiver imposes a cost of misreporting (where the ideal report is defined as $a^S(t)$) on the Sender. Ottaviani and Squintani (2004) also discuss other interesting forms and implications of non-strategic behavior from either of the players.

References

Anderson, R., and W. R. Zame (2000): "Genericity with Infinitely Many Parameters," *Advances in Theoretical Economics*, 1, 1–62.

Austen-Smith, D. (1993): "Interested Experts and Policy Advice: Multiple Referrals under Open Rule," *Games and Economic Behavior*, 5, 3–44.

Austen-Smith, D., and J. S. Banks (2000): "Cheap Talk and Burned Money," *Journal of Economic Theory*, 91(1), 1–16.

Banks, J. S. (1991): *Signaling Games in Political Science*, vol. 46 of *Fundamentals of Pure and Applied Economics*. Harwood Academic Publishers.

Banks, J. S., and J. Sobel (1987): "Equilibrium Selection in Signaling Games," *Econometrica*, 55(3), 647–661.

Battaglini, M. (2002): "Multiple Referrals and Multidimensional Cheap Talk," *Econometrica*, 70(4), 1379–1401.

Bernheim, B. D., and S. Severinov (2003): "Bequests as Signals: An Explanation for the Equal Division Puzzle," *Journal of Political Economy*, 111(4), 733–764.

Birkhoff, G., and G.-C. Rota (1989): *Ordinary Differential Equations*. Wiley, 4th edn.

Blume, A. (1994): "Equilibrium Refinements in Sender-Receiver Games," *Journal of Economic Theory*, 64, 66–67.

——— (1996): "Neighborhood Stability in Sender-Receiver Games," *Games and Economic Behavior*, 13, 2–25.

Chakraborty, A., and R. Harbaugh (2004): "Ordinal Cheap Talk," mimeo, Indiana University.

Cho, I.-K., and D. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102(2), 179–221.

CHO, I.-K., AND J. SOBEL (1990): "Strategic Stability and Uniqueness in Signaling Games," *Journal of Economic Theory*, 50(2), 381–418.

CODDINGTON, E., AND N. LEVINSON (1955): *Theory of Ordinary Differential Equations.* New York: McGraw-Hill.

CRAWFORD, V., AND J. SOBEL (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

DE GARIDEL-THORON, T., AND M. OTTAVIANI (2000): "The Economics of Advice," mimeo, University College London.

DESSEIN, W. (2002): "Authority and Communication in Organizations," *Review of Economic Studies*, 69(4), 811–838.

FARRELL, J. (1993): "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior*, 5(4), 514–531.

FARRELL, J., AND R. GIBBONS (1989a): "Cheap Talk Can Matter in Bargaining," *Journal of Economic Theory*, 48(1), 221–237.

——— (1989b): "Cheap Talk with Two Audiences," *American Economic Review*, 79(5), 1214–1223.

GILLIGAN, T., AND K. KREHBIEL (1987): "Collective Decision-Making and Standing Committees: A Collective Choice Rationale for Restrictive Amendment Procedures," *Journal of Law, Economics, and Organization*, 3, 287–335.

——— (1989): "Assymetric Information and Legislative Rules with a Heterogenous Committee," *American Journal of Political Science*, 33(2), 459–490.

GROSSMAN, S. J., AND M. PERRY (1986): "Perfect Sequential Equilibrium," *Journal of Economic Theory*, 39, 97–119.

HARRIS, M., AND A. RAVIV (2002): "Allocation of Decision-Making Authority," mimeo, Univerisity of Chicago.

JACKSON, M., L. SIMON, J. SWINKELS, AND W. ZAME (2002): "Communication and Equilibrium in Discontinuous Games of Incomplete Information," *Econometrica*, 70(5), 1711–1740.

KOHLBERG, E., AND J.-F. MERTENS (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54(5), 1003–1037.

Kreps, D., and R. Wilson (1982): "Sequential Equilibia," *Econometrica*, 50(4), 863–894.

Krishna, V., and J. Morgan (2001a): "Assymetric Information and Legislative Rules: Some amendments," *American Political Science Review*, 95(2), 435–452.

——— (2001b): "A Model of Expertise," *Quarterly Journal of Economics*, 116(2), 747–775.

——— (2004): "The Art of Conversation: Eliciting Information from Experts through Multi-stage Communication," *Journal of Economic Theory*, 117(2), 147–179.

Levy, G., and R. Razin (2004): "Multidimensional Cheap Talk," mimeo, New York University.

Li, M. (2004): "To Disclose or Not to Disclose: Cheap Talk with Uncertain Biases," mimeo, Concardia University.

Mailath, G. (1987): "Incentive Compatability in Signaling Games with a Continuum of Types," *Econometrica*, 55(6), 1349–1365.

Manelli, A. (1996): "Cheap Talk and Sequential Equilibria in Signaling Games," *Econometrica*, 69(4), 917–942.

——— (1997): "The Never-a-Weak-Best-Response Test in Infinite Signaling Games," *Journal of Economic Theory*, 74, 152–173.

Matthews, S. A. (1989): "Veto Threats: Rhetoric in a Bargaining Game," *Quarterly Journal of Economics*, 104(2), 347–369.

Matthews, S. A., M. Okuno-Fujiwara, and A. Postlewaite (1991): "Refining Cheap-Talk Equilibria," *Journal of Economic Theory*, 55(2), 247–273.

Morgan, J., and P. Stocken (2003): "An Analysis of Stock Recommendations," *RAND Journal of Economics*, 34(1), 183–203.

Morris, S. (2001): "Political Correctness," *Journal of Political Economy*, 109(2), 231–265.

Moscarini, G. (2004): "Competence Implies Credibility," mimeo, Yale University.

Ottaviani, M., and F. Squintani (2004): "Non-Fully Strategic Information Transmission," mimeo, London Business School.

RABIN, M. (1990): "Communiation between Rational Agents," *Journal of Economic Theory*, 51(1), 144–170.

RAMEY, G. (1996): "D1 Signaling Equilibria with Multiple Signals and a Continuum of Types," *Journal of Economic Theory*, 69(2), 508–531.

SOBEL, J. (1985): "A Theory of Credibility," *Review of Economic Studies*, 52(4), 557–573.

STEIN, J. C. (1989): "Cheap Talk and the Fed - A Theory of Imprecise Policy Announcements," *American Economic Review*, 79(1), 32–42.